

Optimal Estimator-Detector Receivers for Space-Time Block Coding

by Olivier Roy

A thesis submitted in partial fulfillment of the requirements for
the Master's degree in Communication Systems

Communication Systems Division (SSC)
School of Computer and Communication Sciences (I&C)
Swiss Federal Institute of Technology Lausanne (EPFL)



March 2004

This work was done under the supervision of

Dr Sylvie Perreau & Prof. Alex Grant

Institute for Telecommunications Research (ITR)
University of South Australia (UniSA)
Mawson Lakes Boulevard
Mawson Lakes 5095
South Australia

and

Prof. Emre Telatar

Information Theory Laboratory (LTHI)
Communication Systems Division (SSC)
School of Computer and Communications Sciences (I&C)
Swiss Federal Institute of Technology Lausanne (EPFL)
1015 Lausanne
Switzerland

Abstract

Most space-time coding schemes can be classified either as non-coherent (decoding is performed without forming an explicit channel estimate) or coherent (decoding is performed conditioned upon a channel estimate as if it were the actual channel realisation). In this thesis, we consider a correlated quasi-static Rayleigh flat fading channel with additive white Gaussian noise. We prove that optimal non-coherent decoding can always be decomposed into a channel estimation step followed by a coherent decoding step. Surprisingly, the required estimators of this coherent approach do not in general minimise the mean square error between the estimated and actual channel realisation. We investigate the characteristics of these estimators and discuss their optimality.

Acknowledgments

First, I would sincerely like to thank my supervisors Dr. Sylvie Perreau and Prof. Alex Grant for the outstanding support they have provided me with throughout my work at the Institute for Telecommunications Research. Their kindness, enthusiasm and knowledge have grandly contributed to the quality this thesis might possess. It has been a real pleasure working with them.

I also wish to thank my supervisor at the Swiss Federal Institute of Technology, Prof. Emre Telatar, for valuable feedbacks and comments about my work here at the University of South Australia.

I would like to express my gratitude to the people I have had valuable discussions with and more generally to all the staff of ITR for ensuring a pleasant and enjoyable workplace.

Finally, I am grateful to my parents for their unconditional support and encouragement throughout all my studies.

Contents

Abstract	v
Acknowledgments	vii
Notation	xiii
List of Acronyms	xv
List of Figures	xvii
List of Tables	xix
1 Introduction	1
2 Space-Time Processing	5
2.1 Multi-Antenna Systems	6
2.1.1 Introduction	6
2.1.2 Fading Channels	7
2.1.3 Diversity	9
2.1.4 From Smart Antennas to MIMO Systems	11
2.1.5 Capacity of Rayleigh Flat Fading MIMO Channels	12
2.2 Space-Time Coding	15
2.2.1 Introduction	15
2.2.2 Design Criteria for Space-Time Codes	16
2.2.3 Space-Time Trellis Codes	18
2.2.4 Space-Time Block Codes	18
2.2.5 An Example: Unitary Space-Time Modulation	19

2.3	Coherent and Non-Coherent Decoding	21
2.3.1	Introduction	21
2.3.2	Channel Estimation Techniques	21
2.3.3	The Need for Channel Estimation	22
3	Optimal Estimator-Detector Receivers	25
3.1	System Model	25
3.2	Optimal Non-Coherent Decoding	26
3.2.1	Optimality in Correlated Fading	27
3.2.2	Optimality in I.I.D. Fading	28
3.2.3	Observations and Discussion	30
3.3	Channel Estimation in Correlated Fading	30
3.3.1	Introduction	31
3.3.2	A Class of Channel Estimators	31
3.3.3	Some Common Channel Estimators	32
3.3.4	Channel Estimation Mean Square Error	35
3.4	Estimator-Detector Receivers	37
3.5	Minimum Codeword Error Probability Estimators	40
3.5.1	Generalised Likelihood Ratio Test	40
3.5.2	Optimal Estimation-Detection Decomposition	42
3.5.3	Optimal Estimator in I.I.D. Fading	49
4	Conclusion	53
A	Matrix Operations	55
A.1	Trace	55
A.2	Kronecker Product and the vec Operator	56
A.3	Block Matrices	57
B	Proofs	59
B.1	Proof of Lemma 3.3.1	59
B.2	Proof of Theorem 3.5.1	60
B.3	Proof of Lemma 3.5.4	61
B.4	Proof of Lemma 3.5.5	63

CONTENTS

xi

B.5 Proof of Theorem 3.5.2 63

B.6 Proof of Theorem 3.5.3 67

Bibliography

69

Notation

Symbols

\mathbb{N}	Non-negative integers
\mathbb{Z}	Integers
\mathbb{C}	Complex numbers
$\mathbb{C}^{m \times n}$	Complex matrices of size $m \times n$
a	A vector or a scalar a (depending on the context)
A	A matrix A
$a_{i,j}$	The element at the i -th row and the j -th column of a matrix A
a_i	The i -th column vector of a matrix A
$A_{i,j}$	The submatrix at position (i, j) of a block matrix A
I_n	The identity matrix of size $n \times n$
$O_{n \times m}$	The all-zero matrix of size $n \times m$
P	The general covariance matrix
P_t	The transmit covariance matrix
P_r	The receive covariance matrix
X	The transmitted codeword
Y	The received matrix
H	The channel matrix
N	The noise matrix
γ	The decision threshold
$\mathcal{CN}(\mu, \Lambda)$	Circularly symmetric complex Gaussian distribution with mean vector μ and covariance matrix Λ
$\text{diag}(d_1, d_2, \dots, d_n)$	A diagonal matrix of size $n \times n$ whose diagonal coefficients are d_1, d_2, \dots, d_n

Operators and Functions

a^T	Transpose of vector a
A^T	Transpose of matrix A

a^*	Conjugate of scalar a or Hermitian transpose of vector a
A^*	Hermitian transpose of matrix A
$\ a\ ^2$	The squared Frobenius norm of vector a
$\ A\ ^2$	The squared Frobenius norm of matrix A
$E[A]$	Expectation of matrix A
$\text{tr}(A)$	Trace of matrix A
$\text{etr}(A)$	Exponential of the trace of matrix A
$ A $	Determinant of matrix A or cardinality of a set A
\otimes	Kronecker product
$\text{vec}(A)$	A vector obtained by stacking the columns of matrix A below one another
a^+	For a scalar a , defined as $\max(a, 0)$

List of Acronyms

AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
CEP	Channel Estimator Parameters
CER	Codeword Error Rate
CNR	Carrier-to-Noise Ratio
CSI	Channel State Information
EDR	Estimator-Detector Receiver
GLRT	Generalised Likelihood Ratio Test
ISI	Intersymbol Interference
MAI	Multiple Access Interference
MAP	Maximum A Posteriori
MCEP	Minimum Codeword Error Probability
MIMO	Multiple Input Multiple Output
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MRC	Maximum Ratio Combining
MSE	Mean Square Error
PSAM	Pilot Symbol Assisted Modulation
PSP	Per-Survivor Processing
SER	Symbol Error Rate
SISO	Single Input Single Output
SNR	Signal-to-Noise Ratio
STBC	Space-Time Block Codes
STC	Space-Time Codes
STTC	Space-Time Trellis Codes
SVD	Singular Value Decomposition
TCM	Trellis Coded Modulation
UMP	Uniformly Most Powerful
USTM	Unitary Space-Time Modulation
ZF	Zero Forcing

List of Figures

2.1	Multipath propagation of a transmitted signal.	7
2.2	A t transmit r receive antenna MIMO system.	12
2.3	Block diagram of a space-time coding system.	16
2.4	Schematic representation of trained and blind estimation. . . .	22
3.1	Block diagram of the ML decoding process.	28
3.2	ZF channel estimation mean square error in i.i.d. and correlated fading.	37
3.3	MMSE channel estimation mean square error in i.i.d. and correlated fading.	38
3.4	Block diagram of an estimator-detector receiver.	39
3.5	Estimation-detection CER for the ZF, MMSE and MCEP estimators in i.i.d. fading.	46
3.6	Estimation-detection CER for the ZF, MMSE and MCEP estimators in both transmit and receive correlated fading. . . .	47
3.7	ZF, MMSE and MCEP channel estimation mean square error in i.i.d. fading.	48
3.8	ZF, MMSE and MCEP channel estimation mean square error in both transmit and receive correlated fading.	49
3.9	Channel estimation mean square error for different values of n . .	51

List of Tables

3.1	Summary of GLRT performance.	42
-----	--------------------------------------	----

Chapter 1

Introduction

Since the appearance of seminal works on space-time information theory and coding [1, 2, 3, 4, 5] two main philosophies have emerged for the design of codes and associated decoding algorithms.

The first strategy takes the view that fundamentally, the coefficients of the space-time channel are unknown and information theoretic principles would tell us to design codes directly for the channel with unknown coefficients but known statistics. This is the non-coherent approach taken in [6, 7, 8, 9] and related works. Here the goal is to design codes and decoders which minimise the decoding error probability, without the possibly unnecessary restrictions due to the use of training sequences (which constrains code design) and channel estimation (which constrains decoder design). Within this framework, optimal detection, in terms of minimising the decoder error probability is the non-coherent maximum likelihood (ML) rule given in [7]. Unless otherwise specified, this is what we shall mean by ML or optimal decoding. Code design for this approach is hard since it is difficult to obtain an expression for the probability of error in closed-form and even more complex to find design rules to minimise it. We shall refer to this first class of strategies as *non-coherent*.

The second strategy is to design the system such that the receiver can easily form some kind of estimate of the fading channel coefficients, which is

subsequently used within a coherent metric as if it were in fact the actual channel realisation [3, 10, 11]. Such sub-optimal detection schemes combine well-known channel estimation techniques with low-complexity decoding rules based on the invalid assumption of perfect channel knowledge. This assumption is usually justified through the use of training sequences or pilot symbols. The interest in this channel estimation based approach is largely motivated by the fact that the optimality provided by ML decoding often comes at the cost of prohibitive complexity. Although systems based on the principles of channel estimation appear to be attractive from a practical implementation point of view, the use of training sequences can only ever have a negative impact upon spectral efficiency. We shall refer to this second class of strategies as *coherent*. We emphasise that in this coherent case, the channel is not a-priori perfectly known, but is always estimated somehow from the received signals.

In this thesis, we wish to compare these two approaches. We shall take the view that at the onset of decoding, the space-time channel gains are unknown. We do assume perfect knowledge of the second order statistics of the channel (i.e. the covariance of the channel gains). We shall regard any training sequences transmitted for the purpose of channel estimation to be part of the coded transmission (i.e. a deterministic prefix for each codeword).

In order to compare coherent and optimal non-coherent detection, we use an estimation-detection approach where decoding is performed as a two step process. First, a set of channel estimates is formed, one for each possible transmitted codeword. Secondly, a coherent metric is computed for each codeword, conditioned on its corresponding channel estimate. The decoder outputs the codeword with the best metric. Working directly on the decision metrics, we show that under the assumptions of unitary codewords and i.i.d.¹ fading, both zero forcing (ZF) and minimum mean square error (MMSE) channel estimators preserve optimality. We relate this result to

¹independent and identically distributed

the generalised likelihood ratio test (GLRT) [12] and show that if the two previous assumptions do not hold (correlated fading and general codewords), optimality may still be retained but that in this case, the required channel estimators are no longer ZF or MMSE.

The outline of this thesis is as follows.

In Chapter 2, we give an overview of space-time processing. We motivate the use of multi-antenna systems by explaining some fundamental concepts. We introduce multiple input multiple output (MIMO) systems by recalling smart antenna systems and give important related information theoretic results. We describe space-time codes (STC) that aim at achieving the capacities promised by the use of MIMO systems. We discuss two kinds of STC and introduce unitary space-time modulation (USTM) as an illustrative example. Finally, coherent and non-coherent approaches are compared by explaining channel estimation techniques and discussing the need for channel estimation in wireless communication systems.

Chapter 3 presents the main contribution of this dissertation. The system model used throughout this thesis is introduced. Optimal non-coherent decoding is reviewed both under i.i.d. and correlated fading. Channel estimation is explained from an analytical perspective and some common channel estimators are reviewed along with their corresponding channel estimation mean square error. The concept of estimator-detector receiver (EDR) is introduced. The GLRT is presented and optimality of such receivers is discussed. Minimum codeword error probability (MCEP) estimators are finally analysed under i.i.d. fading.

We finally offer some conclusion and future directions of research in Chapter 4.

Chapter 2

Space-Time Processing

The aim of this chapter is to introduce some key concepts and results needed for a good understanding of the background upon which this thesis is developed. Some of the material presented here is not strictly necessary for the comprehension of the main results of this dissertation but is included to show the significant advantage of using multi-antenna systems and associated coding schemes in wireless communications.

In Section 2.1, we concentrate on the description of multi-antenna systems. We motivate the use of such systems in practice and introduce the characteristics of fading channels. We also explain the concept of diversity. We introduce MIMO systems with reference to smart antennas and conclude by giving important information theoretic results. Section 2.2 presents some space-time coding strategies, and their associated design criteria, that aim at realising joint encoding for multiple transmit antennas. We explain the concept of space-time codes and describe two main types of STC. We also introduce unitary space-time modulation as an illustrative example. We compare coherent and non-coherent approaches in Section 2.3. We give a short overview of channel estimation techniques and investigate their tradeoffs. Finally, the need for channel estimation is discussed.

2.1 Multi-Antenna Systems

2.1.1 Introduction

During the last decade, wireless cellular networks have grown at an impressive pace. From simple voice calls and text messages to more bandwidth-consuming services such as broadband wireless Internet access and video conferencing, the need for higher data rates has led to the demand for technologies delivering higher capacities and better link reliability than achieved by current systems. This goal is particularly challenging for systems, such as mobile devices, that are power, bandwidth and complexity limited. Since spectral resources are limited, capacity increase is a primary challenge for current wireless network designs. Dealing with an unfriendly transmission medium due to the presence of noise, multipath propagation, interference and fading is also a major issue. In this sense, techniques that improve spectral efficiency and overcome various channel impairments have made an enormous contribution to the growth of wireless communications.

Pioneering work by Winters [13], Telatar [5] and Foschini and Gans [2] has predicted a significant capacity increase associated with the use of multiple transmit and multiple receive antenna systems. This is under the assumptions that the channel can be *accurately tracked* at the receiver and exhibits rich scattering in order to provide *independent transmission paths* from each transmit antenna to each receive antenna. A key feature of MIMO systems, as we will see later on, is the ability to turn multipath propagation, traditionally seen as a major drawback of wireless transmission, into a benefit. This discovery resulted in an explosion of research activity in the realm of MIMO wireless channels for both single user and multiple user communications. In fact, this technology seems to be one of the recent technical advances with a chance of resolving the traffic capacity bottleneck in future Internet-intensive wireless networks. It is surprising to see that just a few years after its invention, MIMO technology already seems poised to be integrated in large-scale

commercial wireless products and applications¹.

2.1.2 Fading Channels

Wireless communication is particularly challenging since it suffers from many channel impairments due to the physical environment characterising the channel (e.g. obstacles on the propagation path) and to interfering signals (e.g. multiple user access). In such an environment, a transmitted signal undergoes different propagation paths with different attenuations, phase shifts and distortions due to the reflecting objects encountered along the path. This phenomenon is shown in Figure 2.1. At the receiver, constructive or



Figure 2.1: Multipath propagation of a transmitted signal.

destructive interference might occur resulting in *multipath fading*. Furthermore, transmitter and/or receiver mobility introduces a time-varying nature to this phenomenon which makes the communication even more challenging.

An effective technique to mitigate multipath fading is known as *transmitter power control* [14]. The idea is that the transmitter pre-distorts the signal before transmission in order to overcome the effect of the channel at the receiver. One major drawback of this method is that the transmitter must be aware of the channel condition as observed at the receiver. In general, this information must be fed back to the transmitter which results in performance

¹See for example 'http://www.pwcwireless.com/demo/pwc_home.asp'

degradation both in terms of rate and complexity, especially in fast multipath fading. Furthermore, to overcome channel fading, the required transmitter power may not be possible due to power limitations. We will see, further in this chapter, that MIMO systems present an effective approach for solving this problem.

Fading channels can generally be classified according to their time, frequency and space characteristics [15, 16]. The behaviour of a fading channel in time can be one of the following:

Fast fading: the time over which the channel characteristics can be considered constant – known as the *coherence time* – is considerably smaller than the time used to transmit the signal. Therefore, the transmitted data will be unequally altered by the channel and suffer from distortion.

Slow fading: the coherence time of the channel is larger than the time used to transmit the signal. Therefore, the whole signal will be equally affected by the channel and will not suffer from the distortion introduced in the fast fading case. The primary degradation is loss in signal-to-noise ratio (SNR).

Similarly, its frequency behaviour can be classified into two main categories:

Frequency selective fading: the bandwidth over which the channel characteristics can be considered constant – known as the *coherence bandwidth* – is considerably smaller than the bandwidth used by the transmitted signal. Therefore, the spectrum of the transmitted data we wish to transmit will be unequally altered by the channel causing the original signal to be spread in time which might introduce intersymbol interference (ISI).

Frequency non-selective fading: the coherence bandwidth of the channel is larger than the bandwidth of the transmitted signal. Therefore, all signal frequency components are equally affected by the channel and no ISI is present. This channel model is also known as *frequency flat fading* [16].

Finally, the channel is characterised in space according to:

Space selective fading: this selectivity is observed in systems using multiple antennas. Space selectivity occurs when the fading of the transmitted signal depends on the spatial location of the antennas, i.e. when the distance between two antennas is considerably bigger than some *coherence distance*. Therefore the signals transmitted using different antennas will be unequally affected.

Space non-selective fading: the coherence distance of the channel is larger than the distance between any pair of antennas. Therefore, the signals transmitted across different transmit antennas will be equally altered.

The channel model that we are going to consider throughout this thesis is a flat fading model where the channel gains are circularly symmetric complex Gaussian random variables or equivalently, have uniformly distributed phase and Rayleigh distributed magnitude. This model is referred to as *Rayleigh flat fading*.

2.1.3 Diversity

As seen in the previous section, channel impairments can be overcome by using transmitter power control. Nevertheless, this method presents major drawbacks that makes it difficult to implement in multipath fast fading communication systems. Another way to combat channel multipath fading is to introduce redundancy into the system by means of *diversity*. Diversity schemes attempt to mitigate the effect of multipath fading by providing the receiver with multiple independently faded replicas of the transmitted signal. The basic idea being that each replica will undergo different independent fades and thus the probability that destructive interference occurs on all the replicas is much smaller than on only one transmitted signal. This will allow the receiver to perform better detection. The most common forms of diversity employed in wireless communications are the following:

Time diversity: replicas of the transmitted signal are provided across time by means of channel coding and time interleaving strategies. Ideally, the

separation between the replicas exceeds the coherence time in order for the replicas to be independent. Therefore, the channel must provide sufficient variations in time for this form of diversity to be useful.

Frequency diversity: replicas of the transmitted waveform are provided on different frequency carriers. Assuming that the coherence bandwidth of the channel is small compared to the bandwidth of the signal, the replicas will suffer independent fades, providing frequency diversity.

Space diversity: this form of diversity, also known as *antenna diversity*, is provided across different antennas at the receiver. Assuming that the separation between the antennas is larger than the coherence distance of the channel, the different received signals may be considered as independent.

Furthermore, diversity schemes can be classified by looking at whether the diversity is applied at the transmitter (*transmit diversity*) or at the receiver (*receive diversity*). The idea behind transmit diversity is to introduce controlled redundancies at the transmitter which can be exploited at the receiver, a subject which will be covered more in depth in Section 2.2 when dealing with space-time codes. Receive diversity schemes may consist of combining different *diversity branches* in order to maximise the quality of the received signal.

The effectiveness of any diversity technique will rely on the availability of independent faded replicas of the transmitted signal at the receiver. A key concept here is that of *diversity order* which is defined as the number of uncorrelated diversity branches available at the transmitter or the receiver. The performance of the system will depend on the ability to properly combine these diversity branches in order to maximise the signal quality. This will be discussed more in depth in the next section.

2.1.4 From Smart Antennas to MIMO Systems

To combat multipath fading, the use of multiple antennas at one end of the transmission link only (transmitter or receiver) was investigated (see [17] for a short tutorial) in order to achieve space diversity. One of the main reasons being that the time interleaving method used for time diversity results in large delays when the channel is slowly varying. Equivalently, frequency diversity is difficult to obtain when the coherence bandwidth is large.

The primary goal of so-called *smart antenna systems* [17] is to efficiently exploit the extra dimension offered by the use of multiple antennas in order to enhance the overall performance of the wireless network. A key concept in smart antennas is that of *beamforming* by which one increases the average SNR by focussing the energy into desired directions, at either the transmitter or the receiver. On the receiver side, a classical approach is to combine the different diversity branches in order to improve the quality of the received signal. The main options used in current systems are [17]:

Selection diversity: it consists of selecting the branch with the highest instantaneous carrier-to-noise ratio (CNR). *Switching diversity* is a variant in which a selected branch is held until its CNR falls under a certain (possibly adaptive) threshold.

Maximum Ratio Combining: this method selects the optimal weights needed to maximise the combined CNR of the different branches. The diversity gain is directly proportional to the number of branches. Although optimal, MRC is expensive to implement since it requires accurate knowledge of the fading coefficients.

Equal gain combining: it is a simplified version of MRC where unit weights are chosen. Nevertheless, the performance of this method is close to that of MRC.

MIMO systems extend smart antennas since multiple antennas are used both at the transmitter and the receiver as shown in Figure 2.2. This may provide

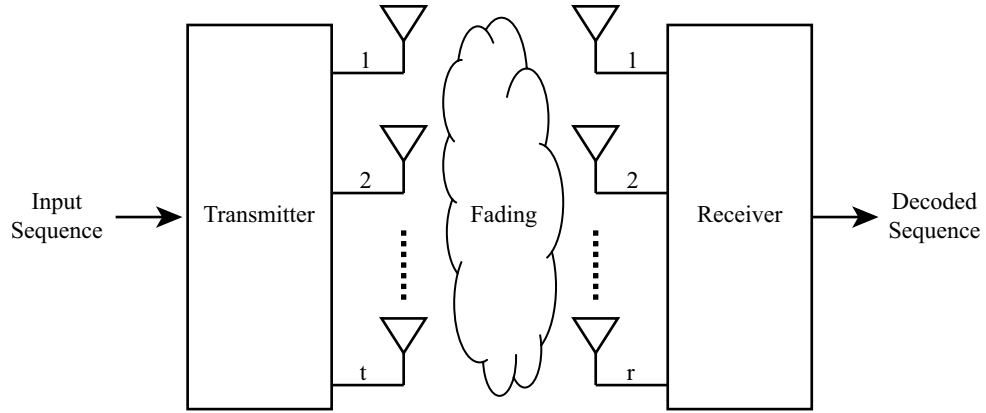


Figure 2.2: A t transmit r receive antenna MIMO system.

additional degrees of freedom in the transmitter or receiver design. A strong analogy can be made between MIMO systems and code-division multiple access (CDMA) transmission. In fact, in CDMA, multiple users access the same time and frequency resources of the channel where each user's data is mixed upon transmission and recovered at reception through its unique code (spreading sequence). In MIMO systems, the independence of the different fading paths provides the input streams with a different *spatial signature* at the receiver since each stream undergoes independent fades. Indeed, the separability of the different MIMO channels relies on a rich scattering environment and therefore MIMO systems are said to *exploit* multipath fading. This is a key difference with smart antennas which try to *mitigate* the effect of multipath fading with the techniques described above.

2.1.5 Capacity of Rayleigh Flat Fading MIMO Channels

In this section, we briefly present some of the fundamental information theoretic results about MIMO systems. We will also consider the main assumptions that are invoked for these results and discuss their validity. An overview of MIMO capacity results for both single and multiuser channels can be found in [18].

Channel capacity is defined as the maximum rate at which data can be transmitted reliably, i.e. with an arbitrarily small probability of error [20]. In his 1948 mathematical theory of communications [19], Shannon derived the capacity of the additive white Gaussian noise (AWGN) channel. It is given by

$$C = W \log \left(1 + \frac{S}{WN_0} \right) \quad (2.1)$$

where W is the bandwidth of the channel in hertz, S the signal power in watts and N_0 is the noise power spectral density in watts per hertz (S/WN_0 is the SNR).

The derivation of capacity for i.i.d. Rayleigh flat fading MIMO channels was independently derived by Telatar [5] and Foschini and Gans [2]. When the channel is constant and perfectly known at both the transmitter and the receiver, it has been shown in [5] that the MIMO channel can be converted to an equivalent set of parallel non-interfering single-input single-output (SISO) channels. The capacity in this case is well known and given by [20]

$$C = \sum_{i=1}^{\min(t,r)} \log \left(\mu \frac{\lambda_i}{\sigma_i^2} \right) \quad (2.2)$$

where t and r are the number of transmit and receive antennas, $\{\sigma_i^2\}$ are the noise variances of the parallel SISO channels, $\{\lambda_i\}$ are the corresponding singular values and μ is chosen such that $\sum_{i=1}^{\min(t,r)} E_i = E$ according to

$$E_i = \left(\mu - \frac{\sigma_i^2}{\lambda_i} \right)^+, \quad (2.3)$$

where E is the total signal power. However, wireless channels are not constant in practice due to the time variations of the propagation environment. Furthermore, perfect knowledge of the propagation coefficients at the transmitter generally involves feedback mechanisms from the receiver. These mechanisms quickly become difficult to implement, especially in fast fading since the channel state information (CSI) becomes rapidly out-of-date.

Thus, the capacity of an ergodic Rayleigh flat fading MIMO channel was derived in [5] and [2] where perfect knowledge of the channel realisation is assumed only on the receiver side. This capacity is given by

$$C = \mathbb{E} \left[\log \left| I_r + \frac{\gamma}{t} H H^* \right| \right] \quad (2.4)$$

where γ denotes the SNR and where the expectation is taken over the random channel realisation H which is assumed to have independent entries. An analytical expression for (2.4) was found in [5] yielding the following important result: if t and r simultaneously become large, the capacity grows linearly with $\min(t, r)$. It is important to recall that two main assumptions have been made for the derivation of the above capacity formula. First, perfect knowledge of the channel gains is assumed at the receiver. Second, the channel coefficients are assumed to be independent over space and time.

For the first assumption to be verified, the receiver can generally estimate the channel response using some widely used channel estimation techniques, such as pilot symbol assisted modulation (PSAM) [21]. We will have a closer look at channel estimation in Section 2.3. For fast fading channels, channel estimation becomes costly and can only have a negative impact on spectral efficiency. Furthermore, channel estimation error also has an undesirable impact on the system performance [22]. Marzetta and Hochwald investigated in [6] the capacity of MIMO systems when the channel coefficients are not known to the receiver. One of the main results they obtained is that making the number of transmit antennas greater than the coherence time does not increase capacity. This is the configuration we are going to use in this thesis since we do not assume perfect knowledge of the channel on the receiver side. This result is somewhat disappointing since it severely limits the achievable rates of fast fading channels, the coherence time being too small to benefit from the deployment of several transmit antennas.

The validity of the second assumption (independence of the channel gains) essentially depends on the nature of the propagation environment. MIMO channels are generally classified as *high/low rank channels*, depending if they

are associated with a rich/poor scattering environment [23]. It has been shown in [24] that the predicted linear growth in capacity becomes less realistic in poor scattering environment. The work in [6] has been extended by Jafar and Goldsmith [25] to the case of correlated fading. It is shown that adding transmit antennas beyond the coherence time is still beneficial, in terms of capacity increase, as long as the channel fading coefficients of the added transmit antennas are spatially correlated with the other transmit antennas. It is interesting to see that transmit correlation fading can be beneficial in the case where only the channel statistics are known. This is obviously not the case when the receiver is aware of the channel realisation [24] since it decreases the diversity gain of the overall MIMO channel. This suggests that depending on the degree of channel knowledge at each end of the transmission link, the problem of code, transmitter and receiver design might change radically.

2.2 Space-Time Coding

2.2.1 Introduction

The information theoretic analysis followed in Section 2.1.5 only provides an upper bound on the achievable rate which might be realised only by codes with very large complexity or latency. Therefore, the development of codes that allow a reasonable tradeoff between performance and complexity is needed to realise the gains offered by MIMO systems in practice.

Space-time codes are codes that exploit both the spatial diversity offered by the use of multiple antennas and the rich temporal structure inherent in digital communication signals and in multipath propagation environments. In other words, STC perform joint coding across space and time to exploit a form of diversity that is referred to as *space-time diversity*. Figure 2.3 shows the block diagram of a space-time coding system. Both on the transmitter and the receiver side, space-time processing can be used to maximise diversity. However, the efficiency of space-time processing at the transmitter may

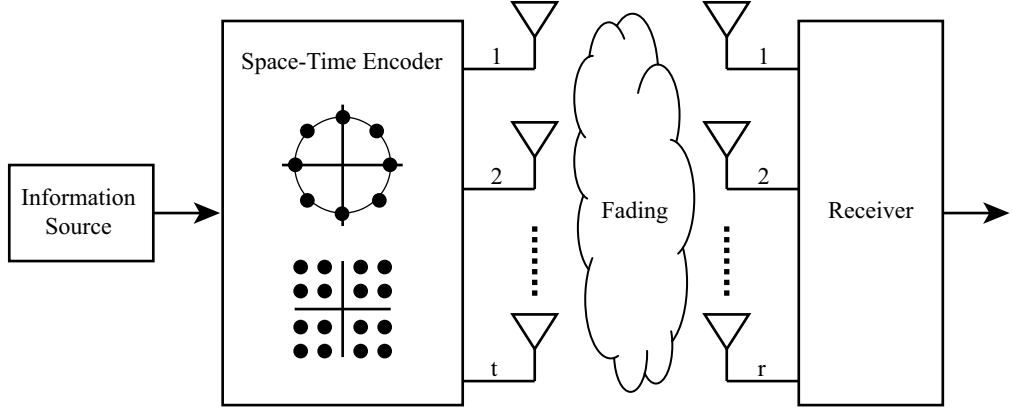


Figure 2.3: Block diagram of a space-time coding system.

be limited by the lack of accurate channel state information.

We will discuss in Section 2.2.3 and 2.2.4 two classes of space-time codes: *space-time trellis codes* (STTC) and *space-time block codes* (STBC).

2.2.2 Design Criteria for Space-Time Codes

The key development of the space-time code concept was revealed in the work of Tarokh, Seshadri and Calderbank [3] where performance criteria for the design of STC were proposed. In this paper, these criteria were applied to the design of space-time trellis codes but can easily be used to evaluate the performance of general space-time codes.

Suppose that for every input symbol k , a space-time encoder generates a sequence of t symbols $c_{k,1}, c_{k,2}, \dots, c_{k,t}$, simultaneously transmitted from t antennas. We define the code vector as $c_k = [c_{k,1}, c_{k,2}, \dots, c_{k,t}]^T$. Suppose that the code vector sequence $C = \{c_1, c_2, \dots, c_l\}$ is transmitted (l consecutive channel uses) and that the decoder decides erroneously in favor of the code vector sequence $\tilde{C} = \{\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_l\}$. We define the *error matrix*

$A \in \mathbb{C}^{t \times t}$ as

$$A(C, \tilde{C}) = \sum_{k=1}^l (c_k - \tilde{c}_k)(c_k - \tilde{c}_k)^*. \quad (2.5)$$

If perfect knowledge of the channel is available at the receiver, the probability of transmitting C and deciding \tilde{C} for a Rayleigh flat fading channel is upper bounded by [3]

$$P(C \rightarrow \tilde{C}) \leq \left(\prod_{k=1}^p \beta_i \right)^{-r} (E_s/4N_0)^{-pr} \quad (2.6)$$

where E_s is the symbol energy, N_0 is the noise power spectral density, p is the rank of the error matrix A and β_i ($i = 1, \dots, p$) are the non-zero eigenvalues of the error matrix A .

Two measures of performance can then be defined based on (2.6). The *coding gain* (or *coding advantage*) achieved by a space-time code is represented by $\prod_{k=1}^p \beta_i$. The term $(E_s/4N_0)^{-pr}$ represents a *diversity gain* (or *diversity advantage*) of pr . The coding gain corresponds to an approximate measure of the gain over an uncoded system operating with the same diversity gain. The diversity advantage is upper bounded by tr since $p \leq t$. Thus, we clearly see from (2.6) that STC should try to maximise both the coding gain and the diversity gain in order to minimise the probability given in (2.6). The two performance measures given above motivate the two following design criteria for space-time codes [3]:

Rank criterion: in order to achieve the maximum diversity tr , the error matrix $A(C, \tilde{C})$ has to be full rank for any pair of code vector sequences C and \tilde{C} . If $A(C, \tilde{C})$ has minimum rank p for some C and \tilde{C} then a diversity of pr is achieved.

Determinant criterion: for a given target diversity gain pr the minimum of the product of the non-zero eigenvalues of $A(C, \tilde{C})$ taken over all pairs of distinct C and \tilde{C} must be maximised in order to maximise the coding gain.

Note that these performance criteria have been shown to be still valid even in the presence of channel estimation error [10].

2.2.3 Space-Time Trellis Codes

Space-Time trellis codes were originally introduced in [3]. They extend trellis coded modulation (TCM) schemes [26] to the case of multiple transmit and receive antennas to combine spatial and temporal diversity techniques. Assuming perfect CSI at the receiver, the authors of [3] proposed STTC that maximise the coding gain for a given diversity advantage. Unfortunately, the corresponding ML decoder requires a relatively high complexity. Therefore, fundamental tradeoffs between rate, diversity, constellation size and trellis complexity were investigated. It was shown that with a constellation of size 2^b , if the diversity advantage is tr , the transmission rate is at most b bits per second per hertz. This was related to the trellis complexity by proving that a multiple antenna system with transmission rate b , employed in conjunction with a space-time trellis code that guarantees a diversity gain p , induces a trellis complexity of at least $2^{b(p-1)}$. Thus, the complexity of STTC decoding increases exponentially as a function of the diversity gain and the transmission rate. Motivated by this major drawback of STTC, the issue of decoding complexity was addressed and led to the space-time block codes discussed in the next section.

2.2.4 Space-Time Block Codes

The field of STBC was initiated by Alamouti [4] who discovered a remarkable scheme for transmission using two transmit antennas. Despite a loss of performance compared to STTC, the impressive gain in complexity obtained by Alamouti's transmit diversity technique still makes it appealing. It is shown in [4] that this scheme provides maximum diversity gain ($2r$) when using two transmit and r receive antennas. This is done by establishing a diversity gain equivalence of this scheme with a $2r$ -branch maximum ratio combining (MRC) scheme [16].

The concept of STBC, where data is encoded using a matrix (a *space-time block code*), was then introduced by Tarokh, Jafarkhani and Calderbank in [27]. Using this new paradigm, Alamouti's scheme was extended to more general code designs. Among them, orthogonal code designs were investigated in [27] where orthogonality is shown to be a key feature in the design of STC to allow linear decoding. In [27], Alamouti's scheme is seen as a special case of orthogonal design.

2.2.5 An Example: Unitary Space-Time Modulation

The use of unitary space-time block codes was introduced by Marzetta and Hochwald in [7] where the corresponding signaling scheme is referred to as *unitary space-time modulation*. Such a modulation scheme is motivated by information theoretic considerations. Assuming that the channel coefficients are i.i.d., it is proved [7] that USTM is nearly optimal (i.e. it achieves a high fraction of the channel capacity) either at high SNR or when the coherence time l is much bigger than the number of transmit antennas t . Furthermore, no perfect knowledge of the channel gains is assumed at the receiver which makes this modulation scheme particularly suitable under very fast fading when it may be impractical to learn the propagation coefficients. Nevertheless, it is shown [7] that USTM can also be justified when the channel coefficients are known to the receiver.

An *unitary space-time signal* is defined as an isotropically distributed matrix $\Phi \in \mathbb{C}^{l \times t}$, i.e. a matrix that obeys $\Phi^* \Phi = I_t$ and whose probability density remains unchanged when left-multiplied by any $l \times l$ unitary matrix. The particular form of unitary space-time signals allows the authors of [7] to get closed-form formulas for both the decoding error probability obtained by an ML receiver in the case the channel is unknown and in the case the channel is known. Based on these results, they note that the two problems are generally very different. Indeed, using two unitary space-time signals Φ_1 and Φ_2 , the optimisation for the first case (channel unknown) is based on the singular values of the product $\Phi_2^* \Phi_1$ whereas for the second case (channel known), it

is based on the singular values of the difference $\Phi_2 - \Phi_1$. Even if an attempt is made to compare the two problems using a Chernoff bound at high SNR, the corresponding ML receivers are shown to be considerably different. Furthermore, when the channel is known, signals are distinguishable that would otherwise be indistinguishable if the channel were unknown. Among there are antipodal pairs of signals as well as signals whose columns are permuted with respect to one another.

The above considerations will be of great interest in this thesis. First because, within our framework, USTM will turn out to have nice properties. Second, the comparison between coherent (channel known) and non-coherent (channel unknown) decoding is the key of this dissertation.

The problem of optimal code design was further investigated by Zhou and Giannakis in [49] in the case of correlated fading. This analysis is performed under the assumption that the channel is perfectly known at the receiver. In the correlated case, getting a closed-form expression for the ML decoding error probability is extremely difficult. Hence, the optimal design is based on an upper bound on the symbol error rate (SER) obtained by an MRC receiver. Based on this bound, they design optimal codes that act as *eigen-beamformers* since energy is distributed along the eigenvectors of the channel covariance matrix. They also discuss the i.i.d. fading case and show that the optimal design uses unitary matrices, leading to the results found in [7] for the ML receiver. Finally, they look at optimal beamforming for MMSE channel estimation purposes and show that the optimal loading is generally not the same as the one used for optimal decoding. This suggests that the optimal beamformer should be chosen differently during the training phase and the transmission phase. Note that this will be of particular interest when we are going to investigate the relationship between channel estimation and optimal detection.

2.3 Coherent and Non-Coherent Decoding

2.3.1 Introduction

As stated in the introduction of this thesis, two approaches to the design of codes and associated decoding algorithms are generally envisioned. The choice of whether to estimate the channel or simply base the decoding process on the channel statistics often depends on several considerations. Complexity, bit error rate (BER) performance, channel characteristics or availability of the requested information are only a few. These tradeoffs are investigated in the two following sections in order to better understand the need for both approaches and therefore to provide a basis on which the rest of this dissertation is developed.

2.3.2 Channel Estimation Techniques

Channel estimation techniques can be broadly classified in two categories: *trained estimation* and *blind estimation*.

The first category, used in most communication systems, consists in extracting the CSI based on some *training sequence*. This training sequence is a sequence of data known at both ends of the transmission link that is used to estimate the channel response at a specific time. Among these techniques, a widely used one is the Pilot Symbol Assisted Modulation (PSAM) method [21]. For time invariant channels, the loss of throughput due to the transmission of training symbols is insignificant. However, the loss of throughput becomes an issue in time varying channels since the CSI must be continuously updated. Furthermore, trained channel estimation techniques are generally not well suited for fast fading channels since the coherence time might not be large enough to allow an accurate channel estimation.

The second category does not rely on the knowledge of any training sequence. Only the received signal is available for the estimation of the channel. Contrary to trained estimation, blind estimation techniques exploit the structure

of the channel and the properties of the input signals in order to perform the estimation. Examples are the probabilistic description of the source or the channel statistics. A survey can be found in [28]. Blind estimation approaches avoid the throughput loss of the trained estimation techniques at the cost of some limited loss in BER performance and more often at the cost of an increased computational complexity.

Trained and blind channel estimation are schematically represented in Figure 2.4. Note that there also exist estimation techniques referred to as *semi-blind* which are based on both methods described above. An example of semi-blind channel estimation can be found in [29].

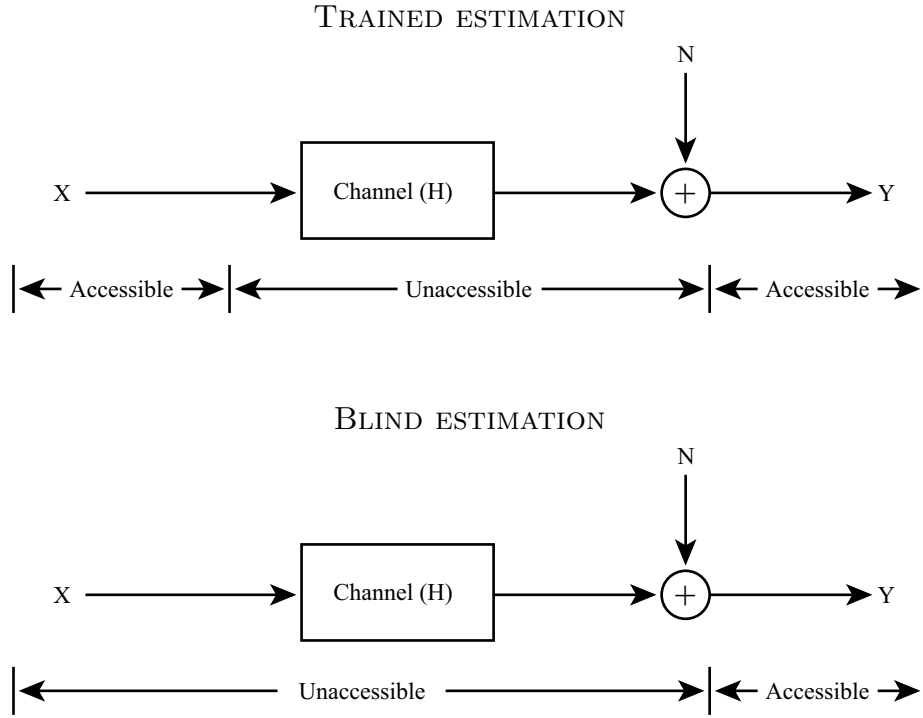


Figure 2.4: Schematic representation of trained and blind estimation.

2.3.3 The Need for Channel Estimation

Estimating the channel realisation or just considering the channel statistics is a fundamental question that arises every time fading is present.

On the one hand, channel estimation forms an essential part of most wireless communication systems. In fact, receiver design plays an important role in the choice of whether or not to estimate the channel. The optimality provided by the ML receiver in terms of minimising the decoding error probability (both for the coherent and non-coherent case) often comes at the cost of prohibitive complexity. Indeed, the ML decoding of a received symbol is done by maximising the a posteriori probability $p(Y|X)$ of the received signal Y over all the possible transmitted symbols in symbol set χ [16]. Thus the ML sequence decoding complexity of a sequence of k symbols will be $\mathcal{O}(|\chi|^k)$ which can be further reduced to $\mathcal{O}(k|\chi|)$ with the use of a Viterbi algorithm [16]. This high computational cost motivates the use of sub-optimal detectors, such as the ZF or the MMSE receivers, that require the computation of a channel estimate. The ZF receiver consists in filtering the received signal in order to suppress the effect of the channel on the symbol sent. This effect could be inter-symbol interference (ISI) in the context of a channel correlated across time or multiple access interference (MAI) in multiuser communications [16]. Nevertheless, the ZF receiver introduces a significant noise enhancement leading to poorer BER performance, especially at low SNR. The MMSE receiver takes the noise into account in the filtering of the received signal yielding better BER performance at the cost of a higher complexity compared to the ZF receiver.

Also, the use of receivers such as the MRC receiver, known for maximising the SNR at the output [16], might be needed (see [49] for example) leading to the need for channel estimation.

The capacity results found in [2] and [5] are based on CSI knowledge at the receiver and thus also motivate a coherent approach to the detection problem.

Finally, the knowledge of the channel might be used efficiently in the design of space-time coding schemes. Scaglione *et al.* proposed in [11] designs for *precoders* and *decoders* that target different optimality criteria and constraints. In this case, the knowledge of CSI can be acquired at the transmitter using some feedback channel.

On the other hand, channel estimation might be difficult. Feedback channels in order to provide the transmitter with CSI is hard to set up in fast varying environments since the estimated channel response becomes rapidly out-of-date. Another important issue that has recently been addressed in many papers is the impact of the channel estimates on the performance of the overall communication system. Zhou and Giannakis investigated in [22] the effect of imperfect channel estimates on adaptive modulation showing how crucial good estimation is. The performance of the MRC receiver was further analysed by Akyildiz and Rao in [30]. Also, the BER performance analysis of single transmit and multiple receive Rayleigh fading channels was carried out in [31] based on noisy predictions. Furthermore, in the context of MIMO systems, the issue of imperfect channel estimates becomes even more important when the number of transmit or receive antennas is large since more coefficients have to be estimated. For the MRC receiver, it was shown in [30] that the quality of the channel estimates degrades with increasing antenna elements. Considering the fact that increasing the number of antennas improves the fading mitigation capabilities of the MRC receiver, it suggests the existence of an optimum number of antenna elements that would take into account the two above opposite effects.

Therefore, choosing to estimate the channel or simply base code design and associated decoding algorithms on channel statistics will depend strongly on the characteristics of the overall transmission link and the information that can be made available to the transmitter and/or the receiver.

Chapter 3

Optimal Estimator-Detector Receivers

3.1 System Model

We will consider a t transmit, r receive space-time channel operating in Rayleigh flat fading environment with l consecutive channel uses ($l \geq t$). The system model admits the following representation [6, 25]:

$$Y = XH + N \quad (3.1)$$

where $Y \in \mathbb{C}^{l \times r}$ is the received matrix and $X \in \mathbb{C}^{l \times t}$ is the transmitted codeword chosen equiprobably from a codebook, $X \in \{X_0, X_1\}$ (this restriction to two codewords is for simplicity only). Note that for ease of notation, we omit the normalisation factor introduced in [6, 25] since it can easily be considered as part of the codeword X . Equation (3.1) can be written element-wise as

$$y_{i,j} = \sum_{k=1}^t x_{i,k} h_{k,j} + n_{i,j} \quad (3.2)$$

for $i = 1, \dots, l$ and $j = 1, \dots, r$. The channel matrix $H \in \mathbb{C}^{t \times r}$ contains the channel gains. In general, the spatially correlated channel can be modeled by [32]

$$\text{vec}(H) = P^{1/2} \text{vec}(H_w) \quad (3.3)$$

where the elements of $H_w \in \mathbb{C}^{t \times r}$ are i.i.d. zero-mean circularly symmetric complex Gaussian random variables with unit variance and $P \in \mathbb{C}^{tr \times tr}$ is the

covariance matrix defined by [32]

$$P \triangleq \mathbb{E}[\text{vec}(H) \text{vec}(H)^*]. \quad (3.4)$$

The correlation model given by (3.3) and (3.4) is capable of representing any correlation effects between the transmit and receive antennas. Nevertheless, it has been shown [33] that an accurate representation of practical spatial correlations (in terms of relative error in capacity) can be described by the following simplified model introduced in [32]:

$$H = P_t^{1/2} H_w P_r^{1/2} \quad (3.5)$$

where $P_t \in \mathbb{C}^{t \times t}$ and $P_r \in \mathbb{C}^{r \times r}$ are the transmit and receive covariance matrices and $H_w \sim \mathcal{CN}(0, I_{tr})$. The channel covariance matrix is given in this case by [32]

$$\mathbb{E}[\text{vec}(H) \text{vec}(H)^*] = P_r \otimes P_t. \quad (3.6)$$

The model described by (3.5) and (3.6) does not take into account the cross-correlation effects between the transmitter and the receiver. This is largely motivated by the fact that this cross-correlation is negligible if the distance between the transmitter and the receiver is large, which is often the case in practice. Therefore (3.5) is the correlation model we are going to use throughout this thesis. Finally, $N \in \mathbb{C}^{l \times r}$ is an additive noise matrix, independent of H , whose elements are i.i.d. zero-mean circularly symmetric complex Gaussian random variables with variance σ^2 , i.e. $N \sim \mathcal{CN}(0, \sigma^2 I_{lr})$.

3.2 Optimal Non-Coherent Decoding

Optimal decoding, in the sense of minimising the decoding error probability, is given by the maximum a posteriori (MAP) rule [16]. When codewords are equiprobable, as it is assumed in our system model, the MAP rule is equivalent to the maximum likelihood (ML) rule. Under the assumption of i.i.d. fading, the latter is well known and widely used in the literature. It is given by [16]

$$\omega_{ML}(Y) \underset{1}{\overset{0}{\gtrless}} |\Omega_0|^r |\Omega_1|^{-r} \quad (3.7)$$

where the decision statistic $\omega_{ML}(Y)$ is defined as

$$\omega_{ML}(Y) \triangleq \text{etr} \left((\Omega_1^{-1} - \Omega_0^{-1}) Y Y^* \right) \quad (3.8)$$

and $\Omega_i = \sigma^2 I_l + X_i X_i^*$ for $i = 0, 1$. However, when fading is correlated, the ML expression is slightly more involved. In the next sections, we introduce the non-coherent ML decoding rule in the correlated case and show how it relates to the i.i.d. case.

3.2.1 Optimality in Correlated Fading

Conditioned upon the received matrix Y according to (3.1), the optimal non-coherent decoding problem is to decide between X_0 and X_1 with minimum probability of error. This is done using the ML decision rule given by [16]

$$p(Y|X_0) \underset{1}{\overset{0}{\gtrless}} p(Y|X_1). \quad (3.9)$$

The covariance matrix of Y is defined as the covariance matrix of the vector $\text{vec}(Y)$ [32] as seen in Section 3.1. Conditioned upon X_i ($i = 0, 1$), the elements of Y are zero-mean circularly symmetric complex Gaussian random variables with covariance matrix

$$\begin{aligned} \Lambda_i &\triangleq \text{E} [\text{vec}(Y) \text{vec}(Y)^* | X_i] \\ &= \text{E} [\text{vec}(X_i H + N) \text{vec}(X_i H + N)^*] \\ &\stackrel{(a)}{=} \text{E} [((I_r \otimes X_i) \text{vec}(H) + \text{vec}(N)) ((I_r \otimes X_i) \text{vec}(H) + \text{vec}(N))^*] \\ &\stackrel{(b)}{=} (I_r \otimes X_i) \text{E} [\text{vec}(H) \text{vec}(H)^*] (I_r \otimes X_i)^* + \text{E} [\text{vec}(N) \text{vec}(N)^*] \\ &\stackrel{(c)}{=} (I_r \otimes X_i) (P_r \otimes P_t) (I_r \otimes X_i)^* + \sigma^2 I_{lr} \\ &\stackrel{(d)}{=} \sigma^2 I_{lr} + P_r \otimes X_i P_t X_i^* \end{aligned} \quad (3.10)$$

using (a)(d) Properties A.2.1, (b) the independence of H and N and (b)(c) the fact that $H \sim \mathcal{CN}(0, P_r \otimes P_t)$ and $N \sim \mathcal{CN}(0, \sigma^2 I_{rl})$. The corresponding conditional densities are [25]

$$p(Y|X_i) = \frac{\text{etr} \left(-\Lambda_i^{-1} \text{vec}(Y) \text{vec}(Y)^* \right)}{\pi^{lr} |\Lambda_i|} \quad \text{for } i = 0, 1. \quad (3.11)$$

Thus, non-coherent ML decoding is done according to the rule

$$\frac{\text{etr} \left(-\Lambda_0^{-1} \text{vec}(Y) \text{vec}(Y)^* \right)}{\pi^{lr} |\Lambda_0|} \underset{1}{\overset{0}{\gtrless}} \frac{\text{etr} \left(-\Lambda_1^{-1} \text{vec}(Y) \text{vec}(Y)^* \right)}{\pi^{lr} |\Lambda_1|} \quad (3.12)$$

which is equivalent to

$$\lambda_{ML}(Y) \underset{1}{\overset{0}{\gtrless}} |\Lambda_0| |\Lambda_1|^{-1} \quad (3.13)$$

where the decision statistic $\lambda_{ML}(Y)$ is defined as

$$\lambda_{ML}(Y) \triangleq \text{etr} \left((\Lambda_1^{-1} - \Lambda_0^{-1}) \text{vec}(Y) \text{vec}(Y)^* \right). \quad (3.14)$$

Figure 3.1 shows a block diagram representing the overall ML decoding process.

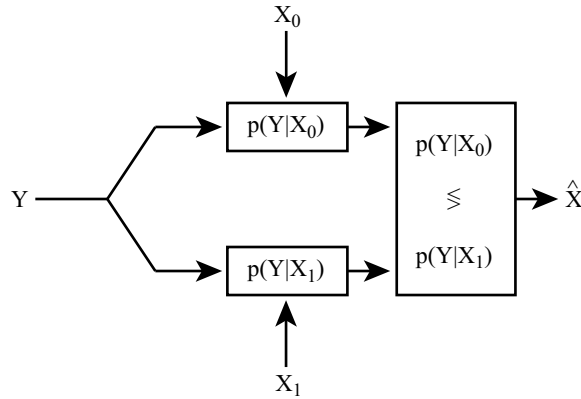


Figure 3.1: Block diagram of the ML decoding process.

3.2.2 Optimality in I.I.D. Fading

The original assumption of i.i.d. channel gains made in [1, 5] in order to prove the capacity increase offered by MIMO systems have led to a large number of publications where the same assumption is made. Therefore, ML decoding is often considered in the context of i.i.d. fading [6, 7]. In this framework, the conditional probability density (3.11) can be further simplified.

In fact, under i.i.d. fading the transmit and receive covariance matrices are

given by $P_t = I_t$ and $P_r = I_r$. Thus, using Properties A.2.1 we can write from (3.10)

$$\begin{aligned}\Lambda_i &= \sigma^2 I_{lr} + P_r \otimes X_i P_t X_i^* \\ &= (I_r \otimes \sigma^2 I_l) + (I_r \otimes X_i X_i^*) \\ &= I_r \otimes (\sigma^2 I_l + X_i X_i^*).\end{aligned}\tag{3.15}$$

Noticing that Λ_i is a block diagonal matrix of size $rl \times rl$ with equal diagonal blocks $(\sigma^2 I_l + X_i X_i^*)$ of size $l \times l$, we can use Lemma A.3.2 to obtain

$$\begin{aligned}|\Lambda_i| &= |I_r \otimes (\sigma^2 I_l + X_i X_i^*)| \\ &= |\sigma^2 I_l + X_i X_i^*|^r \\ &= |\Omega_i|^r.\end{aligned}\tag{3.16}$$

Finally since Λ_i is a positive definite Hermitian matrix, using [34, Th. 7.2.6] and Properties A.2.1 we have

$$\begin{aligned}\text{tr}(\Lambda_i^{-1} \text{vec}(Y) \text{vec}(Y)^*) &= \text{tr}\left((I_r \otimes (\sigma^2 I_l + X_i X_i^*))^{-1} \text{vec}(Y) \text{vec}(Y)^*\right) \\ &= \text{tr}\left((I_r \otimes (\sigma^2 I_l + X_i X_i^*)^{-1}) \text{vec}(Y) \text{vec}(Y)^*\right) \\ &= \text{tr}\left((\sigma^2 I_l + X_i X_i^*)^{-1} Y Y^*\right).\end{aligned}\tag{3.17}$$

Using (3.15), (3.16) and (3.17), the conditional probability density in the case of i.i.d. fading reduces to

$$p(Y|X_i) = \frac{\text{etr}(-\Omega_i^{-1} Y Y^*)}{\pi^{lr} |\Omega_i|^r} \quad \text{for } i = 0, 1\tag{3.18}$$

where

$$\Omega_i = \sigma^2 I_l + X_i X_i^*\tag{3.19}$$

corresponds to the $l \times l$ covariance matrix of the zero-mean circularly symmetric complex Gaussian symbols received at a particular antenna. In this case, optimal decoding is done according to [16],

$$\omega_{ML}(Y) \stackrel{0}{\underset{1}{\gtrless}} |\Omega_0|^r |\Omega_1|^{-r}\tag{3.20}$$

where the decision statistic $\omega_{ML}(Y)$ is defined as

$$\omega_{ML}(Y) \triangleq \text{etr}\left((\Omega_1^{-1} - \Omega_0^{-1}) Y Y^*\right).\tag{3.21}$$

3.2.3 Observations and Discussion

By comparing how ML decoding is performed in the i.i.d. and in the correlated case, we can make the following observations.

Looking at Equation (3.10), we note that if there is only correlation between the transmit antennas ($P_r = I_r$), the simplifications of Section 3.2.2 can be applied and the covariance matrix Ω_i can be replaced by $(\sigma^2 I_l + X_i P_t X_i^*)$. This suggests that dealing with transmit correlation only is in general not a problem since it is equivalent to i.i.d. fading where the codewords X_i are replaced by $X_i P_t^{1/2}$. In other words, the use of a transmit correlation only fading channel is equivalent to an i.i.d. fading channel with a different codebook. Thus, if no particular assumption is made on the codewords, the i.i.d. fading channel results can easily be extended in the presence of transmit correlation only. As we will see in the rest of this thesis, receive correlation is more critical and in general results found for the i.i.d. case are not so trivially extended. It is also important to emphasise that Ω_i does not correspond to the covariance matrix of the received signal Y . This latter is given by the matrix Λ_i of Equation (3.15). Nevertheless, when no receive correlation is present, the columns of Y (corresponding to each receive antenna) are i.i.d. zero-mean circularly symmetric complex Gaussian random vectors with covariance matrix Ω_i . This allows us to perform the simplifications made in Section 3.2.2.

3.3 Channel Estimation in Correlated Fading

Channel estimation was explained from a conceptual point of view in Section 2.3.2. Now we take a closer look at how it is actually performed from an analytical point of view. A convenient representation of the channel estimators under consideration is introduced and some common estimators are presented along with their corresponding mean square error.

3.3.1 Introduction

The problem of channel estimation can be stated as follows [35]: let $Y \in \mathbb{C}^{l \times r}$ be the received matrix corresponding to some transmitted data $X \in \mathbb{C}^{l \times t}$. A linear channel estimate $\hat{H} \in \mathbb{C}^{t \times r}$ is given by

$$\text{vec}(\hat{H}) = K \text{vec}(Y) \quad (3.22)$$

for a matrix $K \in \mathbb{C}^{tr \times lr}$ that may be chosen according to some optimisation criterion (e.g. ZF, MMSE) and that might be a function of the corresponding transmitted data X . This choice aims to give \hat{H} some desired properties such as minimising the channel estimation mean square error in the MMSE case. The matrix K will be referred to as the *channel estimator*. The channel estimation performed using (3.22) is capable of representing any correlation present in the actual channel realisation and therefore in the received matrix Y . In the case of transmit correlation only, some simplifications can be done as seen for the ML decoding rule in Section 3.2.2. In fact, since the columns of H are i.i.d. zero-mean circularly symmetric complex Gaussian random vectors with covariance matrix P_t , estimation can be performed column-wise by a $t \times l$ matrix \tilde{K} . In other words, using Properties A.2.1, Equation (3.22) can be rewritten as

$$\begin{aligned} \text{vec}(\hat{H}) &= (I_r \otimes \tilde{K}) \text{vec}(Y) \\ &= \text{vec}(\tilde{K}Y) \end{aligned} \quad (3.23)$$

or equivalently

$$\hat{H} = \tilde{K}Y \quad (3.24)$$

for some matrix $\tilde{K} \in \mathbb{C}^{t \times l}$. Therefore, when dealing with i.i.d. fading, we will focus on Equation (3.24) instead of (3.22).

3.3.2 A Class of Channel Estimators

Under i.i.d. fading, we now introduce a class of estimators $\tilde{K} \in \mathbb{C}^{t \times l}$ that can be written in terms of singular values. This will allow us to more easily compare the different channel estimators we are interested in.

Assume that the received matrix $Y \in \mathbb{C}^{l \times r}$ used to evaluate the channel response according to (3.24) corresponds to the transmitted signal $X \in \mathbb{C}^{l \times t}$ whose singular value decomposition (SVD) is given by

$$X = U \Sigma V^* \quad (3.25)$$

where $U \in \mathbb{C}^{l \times l}$ and $V \in \mathbb{C}^{t \times t}$ are unitary matrices and where $\Sigma \in \mathbb{C}^{l \times t}$ is a matrix of the form

$$\Sigma = \begin{bmatrix} \bar{\Sigma} \\ O_{(l-t) \times t} \end{bmatrix} \quad (3.26)$$

with $\bar{\Sigma} \in \mathbb{C}^{t \times t}$ being a diagonal matrix whose diagonal elements are the singular values of X given by $\sigma_1, \sigma_2, \dots, \sigma_t$. We now define a class of estimators \tilde{K} as

$$\tilde{K} = V \tilde{D} U^* \quad (3.27)$$

where $\tilde{D} \in \mathbb{C}^{t \times l}$ is a matrix of the form

$$\tilde{D} = \begin{bmatrix} \bar{D} & O_{(l-t) \times t} \end{bmatrix} \quad (3.28)$$

with $\bar{D} \in \mathbb{C}^{t \times t}$ being a diagonal matrix whose diagonal elements $\tilde{d}_1, \tilde{d}_2, \dots, \tilde{d}_t$ are determined by the optimisation criterion chosen and are referred to as the *channel estimator parameters* (CEP).

As we will see, the estimators we are going to consider throughout this thesis belong to this class and can thus be described by their corresponding CEP.

3.3.3 Some Common Channel Estimators

In this section, we analyse more carefully three channel estimators that will be of particular interest in our discussion and that are widely encountered in the literature: the ZF, MMSE and ML estimators.

Zero Forcing Estimator

One of the simplest estimators, in terms of complexity, is given by looking for a matrix $K \in \mathbb{C}^{tr \times lr}$ such that the channel estimate \hat{H} minimises

$$\| \text{vec}(Y) - \text{vec}(X \hat{H}) \|^2 \quad (3.29)$$

given the received matrix Y and the corresponding input data X . The solution of this optimisation problem can be inferred from [36], is given by

$$\begin{aligned} K &= \left((I_r \otimes X)^* (I_r \otimes X) \right)^{-1} (I_r \otimes X)^* \\ &= I_r \otimes (X^* X)^{-1} X^* \end{aligned} \quad (3.30)$$

$$= I_r \otimes \tilde{K} \quad (3.31)$$

and is referred to as the *zero-forcing estimator*. This terminology comes from the fact that this criterion forces the ISI to zero when used in the context of equalisation [16]. We see in (3.30) that neither the noise nor the statistics of the channel are taken into account in the ZF estimator. A poor channel estimate, especially at low SNR, is thus the price to pay for the low complexity provided by this estimator [35]. Since the coefficients of the channel are not assumed to be random, or in other words that the statistics of the channel are neglected, the ZF estimator can be entirely described by the $l \times t$ matrix $\tilde{K} = (X^* X)^{-1} X^*$ as stated in Section 3.3.1. The computation of the matrix \tilde{K} requires $\mathcal{O}(lt^2)$ operations. If X is not full rank, the matrix $(X^* X)$ is not invertible and the ZF estimator cannot be computed. This is a problem that the MMSE estimator avoids as we will see in the next section. Finally, using the SVD of X , the matrix \tilde{K} can be written as

$$\begin{aligned} \tilde{K} &= (X^* X)^{-1} X^* \\ &= (V \Sigma^* U^* U \Sigma V^*)^{-1} V \Sigma^* U^* \\ &= V (\Sigma^* \Sigma)^{-1} V^* V \Sigma^* U^* \\ &= V (\Sigma^* \Sigma)^{-1} \Sigma^* U^* \end{aligned}$$

and thus admits the following CEP

$$\tilde{d}_j = \frac{\sigma_j^*}{|\sigma_j|^2} \quad (3.32)$$

for $j = 1, \dots, t$.

Minimum Mean Square Error Estimator

The MMSE estimator, as its name implies, is obtained by finding the matrix $K \in \mathbb{C}^{tr \times lr}$ such that the mean square error between the channel estimate \hat{H}

and the actual channel realisation H is minimised. Finding the minimum of the quantity

$$\mathbb{E} \left[\|\text{vec}(H) - \text{vec}(\hat{H})\|^2 \right] \quad (3.33)$$

is obtained by choosing K as [36]

$$\begin{aligned} K &= (P_r \otimes P_t) (I_r \otimes X)^* (\sigma^2 I_{rl} + (I_r \otimes X) (P_r \otimes P_t) (I_r \otimes X)^*)^{-1} \\ &= (P_r \otimes P_t X^*) (\sigma^2 I_{rl} + P_r \otimes X P_t X^*)^{-1}. \end{aligned} \quad (3.34)$$

If there is no receive correlation ($P_r = I_r$), the estimator K reduces to

$$\begin{aligned} K &= I_r \otimes P_t X^* (\sigma^2 I_l + X P_t X^*)^{-1} \\ &= I_r \otimes \tilde{K} \end{aligned} \quad (3.35)$$

where the familiar $t \times l$ matrix $\tilde{K} = P_t X^* (\sigma^2 I_l + X P_t X^*)^{-1}$ is widely used in the literature [36]. The major advantage of the MMSE estimator is that the noise and the channel statistics are taken into account in the estimation process. This reduces the degradation of the estimates at low SNR. Furthermore, X does not have to be full rank for the MMSE estimate to be computed. Nevertheless, the complexity of this estimator is higher than the ZF since it involves the inversion of the $rl \times rl$ matrix of (3.34) instead of the $t \times t$ matrix of (3.30). The number of channel uses l being in general much bigger than the number of transmit antennas t , we clearly see the increase of complexity compared to the ZF estimator. Note that in the case of transmit correlation only, this problem can be solved by the use of the matrix inversion lemma [34, Sec. 0.7.4]. Indeed, the inversion of the $l \times l$ matrix (3.35) can be reduced to the inversion of a $t \times t$ matrix. It would be interesting to see if a similar lemma could be stated for a matrix inversion of the form of Equation (3.34). Assuming i.i.d. fading ($P_t = I_t$, $P_r = I_r$) and using the SVD of X , the matrix \tilde{K} can be written as

$$\begin{aligned} \tilde{K} &= X^* (\sigma^2 I_l + X X^*)^{-1} \\ &= V \Sigma^* U^* (\sigma^2 U U^* + U \Sigma V^* V \Sigma^* U^*)^{-1} \\ &= V \Sigma^* U^* U (\sigma^2 I_l + \Sigma \Sigma^*)^{-1} U^* \\ &= V \Sigma^* (\sigma^2 I_l + \Sigma \Sigma^*)^{-1} U^* \end{aligned}$$

and thus admits the following CEP

$$\tilde{d}_j = \frac{\sigma_j^*}{|\sigma_j|^2 + \sigma^2} \quad (3.36)$$

or equivalently

$$\tilde{d}_j = \begin{cases} \frac{\sigma_j^*}{|\sigma_j|^2} \left(1 - \frac{\sigma^2}{|\sigma_j|^2 + \sigma^2}\right) & \text{if } \sigma_j \neq 0 \\ 0 & \text{if } \sigma_j = 0 \end{cases} \quad (3.37)$$

for $j = 1, \dots, t$.

Maximum Likelihood Estimator

The ML estimator will be of particular interest when introducing the GLRT. Here the channel estimate \hat{H} is found by maximising the likelihood $p(Y|X, \hat{H})$ over all possible values of \hat{H} given the received matrix Y and the corresponding input data X . Since this density is given under our Gaussian assumptions by [7]

$$p(Y|X, \hat{H}) = \frac{1}{(\sigma^2 \pi)^{lr}} \text{etr} \left(-\frac{1}{\sigma^2} (Y - X\hat{H}) (Y - X\hat{H})^* \right), \quad (3.38)$$

it can be seen using Properties A.2.1 that maximising $p(Y|X, \hat{H})$ is equivalent to minimising $\|\text{vec}(Y) - \text{vec}(X\hat{H})\|^2$ which corresponds to the ZF criterion (3.29). In our framework, the ZF and ML estimators are thus equivalent.

3.3.4 Channel Estimation Mean Square Error

A common measure to quantify the quality of a channel estimate is given by its mean square error (MSE) [16]. Under i.i.d. fading, the analytical expression of the MSE can be easily found in the literature [49]. In the following lemma, we derive the MSE under the correlation model adopted in this dissertation.

Lemma 3.3.1 *Let $\hat{H} \in \mathbb{C}^{t \times r}$ be a channel estimate computed as*

$$\text{vec}(\hat{H}) = K \text{vec}(Y)$$

for some matrix $K \in \mathbb{C}^{tr \times lr}$. The channel estimate mean square error is given by

$$\begin{aligned} & \mathbb{E} \left[\|\text{vec}(H) - \text{vec}(\hat{H})\|^2 | X \right] = \\ & \text{tr} \left((I_{tr} - K(I_r \otimes X))^* (I_{tr} - K(I_r \otimes X)) (P_r \otimes P_t) + \sigma^2 K^* K \right). \end{aligned}$$

Proof: See Appendix B.1. □

From Lemma 3.3.1, it is straightforward to compute the estimation mean square error obtained for the ZF estimator, denoted ϵ_{ZF} , by replacing K given by Equation (3.30) to get

$$\epsilon_{ZF} = r\sigma^2 \text{tr}((X^* X)^{-1}). \quad (3.39)$$

We see from (3.39) that ϵ_{ZF} does not depend on the channel statistics and is therefore invariant under i.i.d. and correlated fading. It also increases linearly with the number of receive antennas and the noise variance. We plot in Figure 3.2 the channel estimation MSE obtained for a $t = 3$ transmit and $r = 2$ receive antenna MIMO systems. The training matrix X is chosen arbitrarily among the full rank matrices of size $l \times t$ where $l = 5$.

Similarly, we plot in Figure 3.3 the channel estimation MSE obtained for the MMSE channel estimator. The parameters are the same as before and the transmit and receive covariance matrices are chosen arbitrarily. From Equation (3.34), we see that the matrix K tends to the all zero matrix of size $tr \times lr$ as $\sigma \rightarrow \infty$. Thus looking at Lemma 3.3.1, the behaviour of the curves of Figure 3.3 at low SNR is dictated by $\text{tr}(P_r \otimes P_t)$ which explains the gaps between the curves corresponding to different propagation environments. If the channel gains are highly correlated, the estimation of one coefficient gives us information about the others and thus the channel estimation process is less error-prone. In contrast for the ZF channel estimator, the MSE is upper bounded as the noise becomes large which is a major advantage of the MMSE estimator.

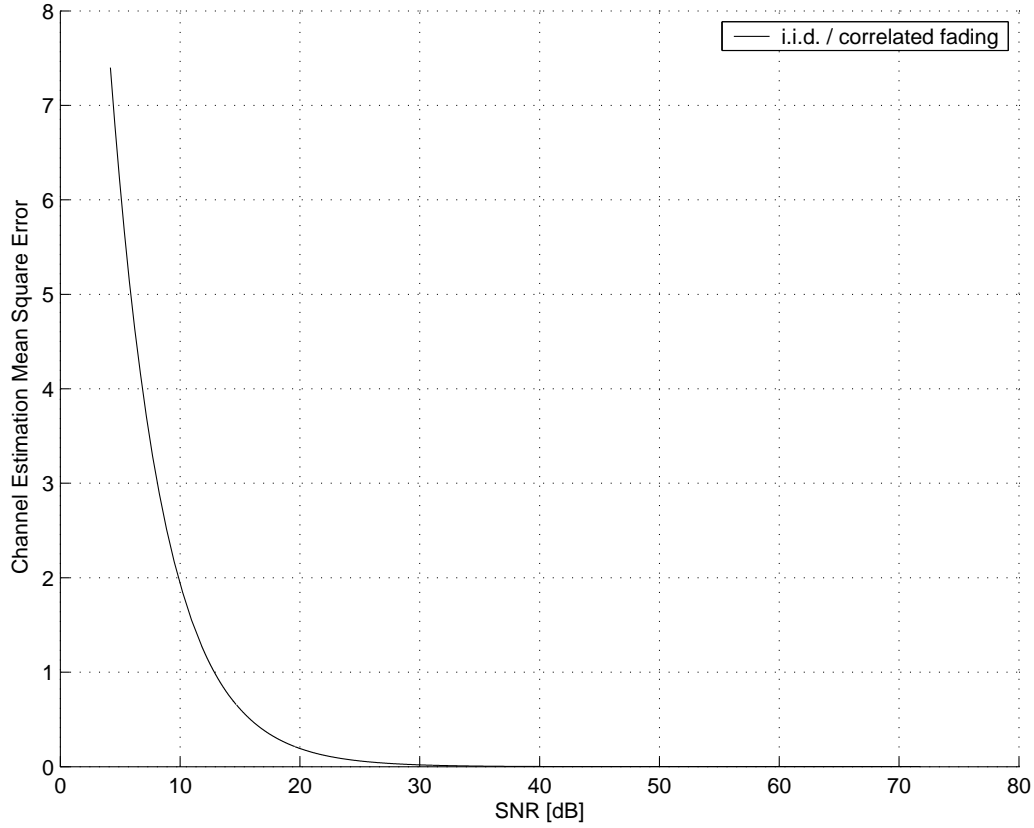


Figure 3.2: ZF channel estimation mean square error in i.i.d. and correlated fading.

3.4 Estimator-Detector Receivers

The non-coherent ML decoding reviewed in Section 3.2 performs optimal decoding in a non-coherent manner, i.e. without the use of channel estimates. In this section, we consider a coherent approach to the originally non-coherent decoding problem described previously. We motivate the approach and introduce the main problem addressed in this thesis.

Consider decoding according to the following, possibly sub-optimal, *two step process*. First, two channel estimates (one for each possible codeword) are

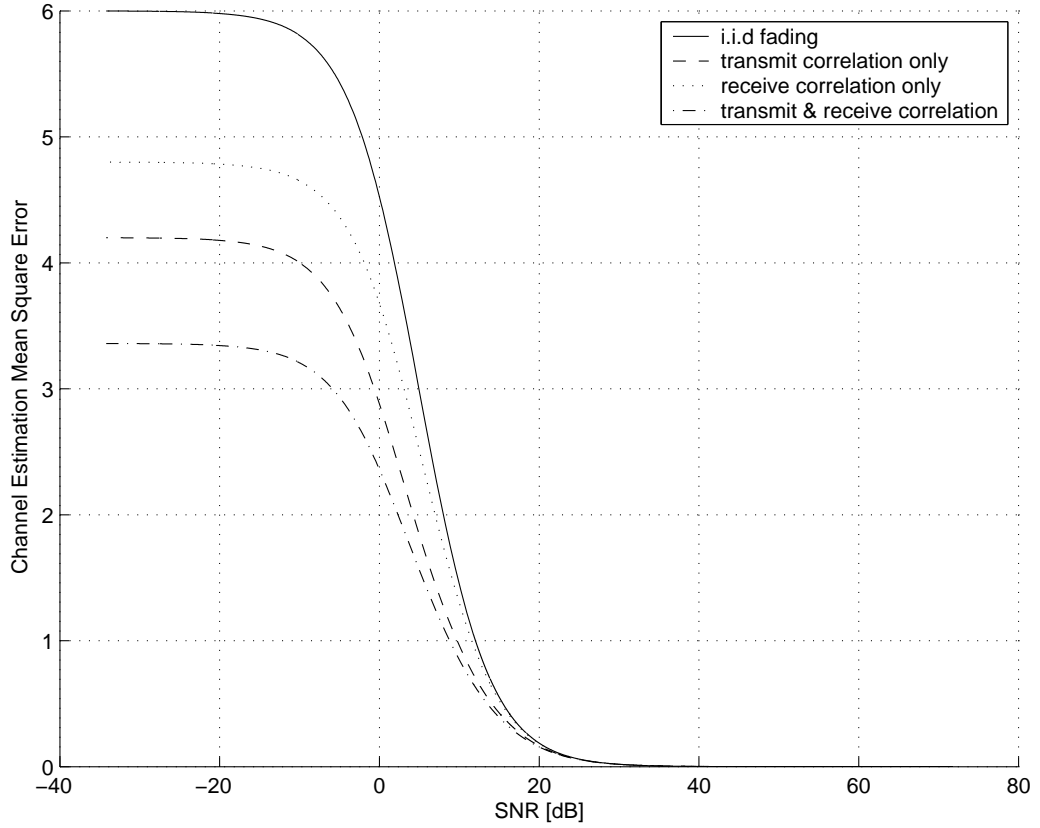


Figure 3.3: MMSE channel estimation mean square error in i.i.d. and correlated fading.

computed as

$$\text{vec}(\hat{H}_0) = K_0 \text{vec}(Y) \quad (3.40)$$

$$\text{vec}(\hat{H}_1) = K_1 \text{vec}(Y) \quad (3.41)$$

where $K_0 \in \mathbb{C}^{tr \times lr}$ and $K_1 \in \mathbb{C}^{tr \times lr}$ may be chosen according to some optimisation criterion as seen in Section 3.3.3. The channel estimator K_i ($i = 0, 1$) is computed under the assumption that the codeword X_i was sent. Now, using these channel estimates, we define a coherent detection approach to the originally non-coherent decoding problem. Let

$$\mu_i(Y) \triangleq \exp \left(\left\| \text{vec}(Y) - \text{vec}(X_i \hat{H}_i) \right\|^2 \right) \quad (3.42)$$

$$= \text{etr} \left(\text{vec}(Y - X_i \hat{H}_i)^* \text{vec}(Y - X_i \hat{H}_i) \right) \quad (3.43)$$

be the coherent metric for X_i , using \hat{H}_i as if it were the true channel realisation. Within this framework, the decision statistic is

$$\psi(Y) = \exp \left(\|\text{vec}(Y) - \text{vec}(X_1 \hat{H}_1)\|^2 - \|\text{vec}(Y) - \text{vec}(X_0 \hat{H}_0)\|^2 \right) \quad (3.44)$$

with decision rule

$$\psi(Y) \underset{1}{\overset{0}{\gtrless}} \gamma \quad (3.45)$$

for some decision threshold γ . This is referred to as an *estimator-detector receiver*. Such a receiver is described by the block diagram given in Figure 3.4. The motivation behind this approach is the following. First, the channel es-

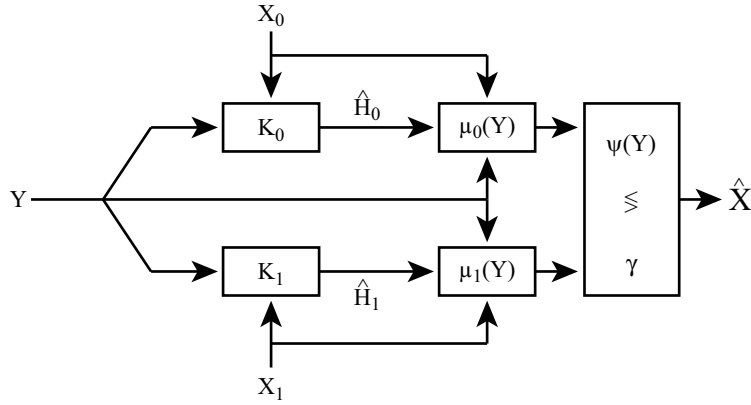


Figure 3.4: Block diagram of an estimator-detector receiver.

timates are computed without the use of any deterministic training scheme such as PSAM techniques [21]. This will allow us to make a fair comparison of this estimation-detection approach to the optimal non-coherent ML decoding since no additional information is needed to compute the channel estimates. Second, the channel estimation approach adopted in this two step process can be seen as a deterministic blind estimation of the channel where the input codeword X_i is part of the unknown parameters (see [28] for details). Finally, as we will see in Section 3.5.1, under our Gaussian assumptions the GLRT is a particular case of an EDR where the matrices K_0 and K_1 are chosen according to the ML criterion.

In general, closed-form expressions for the decoding error probability are

extremely hard to obtain. Therefore, directly finding an EDR that preserves the optimality provided by the ML non-coherent decoder is difficult. Here however, this problem turns out to be easily solved by directly matching the corresponding decision rules. Thus, the rest of this thesis will focus on finding channel estimators that allow the estimation-detection approach given by Equations (3.44) and (3.45) to be equivalent to the optimal decoding given by (3.14) and (3.13).

3.5 Minimum Codeword Error Probability Estimators

This section presents the main contribution of this thesis. We first introduce the GLRT and show how this test relates to our EDR framework. We recall that the GLRT performs optimal detection under the assumptions of both i.i.d. fading and unitary codewords. We prove that the same result holds when the ML estimates used in the GLRT are replaced by MMSE estimates. Then, we extend this result by showing that, even if the two previous assumptions do not hold, the EDR described in Section 3.4 can still preserve optimality by carefully choosing the channel estimators and the corresponding decision threshold. Under i.i.d. fading, we give the CEP of these “optimal” estimators, referred to as the *minimum codeword error probability* estimators, and compare them to the ZF and MMSE estimators.

3.5.1 Generalised Likelihood Ratio Test

The GLRT was first introduced in the context of signal detection [12]. The aim is to design tests that detect the presence of a source signal in a noisy environment. When the detection problem depends on unknown parameters (characterised by the vector θ), it is referred to as a *composite hypothesis* testing problem [12, Sect. 2.5]. In the case θ has a known distribution, the problem reduces to a simple hypothesis testing problem where integration is performed over θ . When θ is not a random variable, the goal is to design tests that perform as well as if the value of θ were known. Such tests are

referred to as *uniformly most powerful* (UMP). Unfortunately, UMP tests do not always exist and alternative tests such as the GLRT are needed.

In our framework, the detection problem consists of deciding between the two hypothesis X_0 and X_1 depending on the unknown parameter H . The GLRT detection is performed according to [12, Sect. 2.5]

$$\frac{\max_H p(Y|X_0, H)}{\max_H p(Y|X_1, H)} \underset{1}{\overset{0}{>}} \gamma \quad (3.46)$$

which can be equivalently rewritten as

$$\frac{p(Y|X_0, \hat{H}_0)}{p(Y|X_1, \hat{H}_1)} \underset{1}{\overset{0}{>}} \gamma \quad (3.47)$$

where the channel estimate under hypothesis i , denoted \hat{H}_i , is chosen as the ML channel estimate given by

$$\hat{H}_i = \arg \max_H p(Y|X_i, H). \quad (3.48)$$

Under our Gaussian assumptions, the decision rule (3.46) is equivalent to the EDR decision rule (3.45) where the channel estimate \hat{H}_i is given by

$$\hat{H}_i = (X_i^* X_i)^{-1} X_i^* Y \quad (3.49)$$

as seen in Section 3.3.3. It is important to recall that in our system model, the above ML channel estimate corresponds to a ZF estimate.

It has been shown [45] that under the assumptions of both unitary signaling and i.i.d. fading, the GLRT performs optimal detection (with $\gamma = 1$) or in other words that the GLRT and the non-coherent ML receivers are equivalent. The following theorem extends this result to MMSE estimates.

Theorem 3.5.1 *Let X_0 and X_1 be unitary codewords and the channel fading be i.i.d. Then, an EDR using MMSE channel estimates and a decision threshold $\gamma = 1$ is optimal.*

Proof: See Appendix B.2. □

Theorem 3.5.1 shows that the particular form of unitary codes allows us some freedom in the design of the estimator that retains optimality since both ZF and MMSE channel estimates preserve optimal detection. Furthermore, good channel estimation is not crucial in minimising the decoding error probability since we can do optimal detection at very low SNR with a ZF channel estimate whose MSE is poor compared to an MMSE estimate. This result can be intuitively understood by considering the symmetry introduced in the detection problem by the use of unitary codewords. By taking the logarithm on both sides of the decision rule (3.45), the threshold becomes 0 and the decision statistic (3.44) can be modified up to some positive multiplicative constant without changing the decision rule and thus without altering the corresponding decoding error probability.

3.5.2 Optimal Estimation-Detection Decomposition

The GLRT described in Section 3.5.1 preserves optimal detection under the assumptions of both unitary signaling and i.i.d. fading. However, this test becomes suboptimal if one of these two assumptions does not hold [45]. Nevertheless, it has been shown in [45] that the non-coherent ML receiver asymptotically (at high SNR) converges to the GLRT receiver for unitary codewords and correlated fading. A summary of GLRT performance is given in Table 3.1. In this section, we address the problem of optimal EDR de-

	<i>I.I.D. Fading</i>	<i>Correlated Fading</i>
<i>Unitary Codewords</i>	Optimal	Suboptimal (Optimal as $\text{SNR} \rightarrow \infty$)
<i>General Codewords</i>	Suboptimal	Suboptimal

Table 3.1: Summary of GLRT performance.

coding when no specific assumption is made, either on the codewords or on the fading. In the optimal estimator-detector receiver theorem, we will prove that under correlated fading we can always find a channel estimator such

that the EDR approach presented in Section 3.4 is optimal for some decision threshold γ . In other words, we show that optimal non-coherent ML decoding can always be decomposed in a coherent manner using estimator-detector receivers. This result leads to the *optimal estimator-detector receivers*.

We first introduce a few results needed for the proof of Theorem 3.5.2.

Definition 3.5.1 (Diagonal Element Matrix) *Let $A \in \mathbb{C}^{mn \times mn}$ be a block matrix of the form*

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,m} \\ A_{2,1} & A_{2,2} & \dots & A_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \dots & A_{m,m} \end{bmatrix}$$

where $A_{i,j} \in \mathbb{C}^{n \times n}$ are diagonal matrices ($i, j = 1, \dots, m$). We will call A a *diagonal element matrix*.

Remark: The term *diagonal element* is chosen in order not to be confused with *block diagonal* or *diagonal block* which both designate a block matrix whose diagonal blocks are general square matrices and whose off-diagonal blocks are the all zero matrices [34, Sec. 0.9.2] (see also Definition A.3.2).

Lemma 3.5.1 *Let $A \in \mathbb{C}^{mn \times mn}$ and $B \in \mathbb{C}^{mn \times mn}$ be diagonal element matrices. Then AB is a diagonal element matrix.*

Proof: The product AB is performed block-wise to get

$$(AB)_{i,j} = \sum_{k=1}^m A_{i,k} B_{k,j}$$

which is a diagonal matrix since $A_{i,j} \in \mathbb{C}^{n \times n}$ and $B_{i,j} \in \mathbb{C}^{n \times n}$ are diagonal ($i, j = 1, \dots, m$). \square

Lemma 3.5.2 *Let $A \in \mathbb{C}^{mn \times mn}$ be a diagonal element matrix. Then for any $k \in \mathbb{N} \setminus \{0\}$, A^k is a diagonal element matrix.*

Proof: The proof follows directly from Lemma 3.5.1. \square

Lemma 3.5.3 *Let $A \in \mathbb{C}^{mn \times mn}$ be a diagonal element matrix. Then for any scalar $k \in \mathbb{Z}$, A^k is a diagonal element matrix (if $k < 0$, A has to be non-singular).*

Proof: If A is a non-singular diagonal element matrix so is A^{-1} since applying recursively the inversion formula given in Lemma A.3.1 involves only (possibly block-wise) additions, multiplications and inversions of diagonal matrices. We then use Lemma 3.5.2 to conclude. Note that A^0 is defined as the identity matrix of size $mn \times mn$ and is thus also a diagonal element matrix. \square

Lemma 3.5.4 *Let $A \in \mathbb{C}^{mn \times mn}$ be a positive definite Hermitian diagonal element matrix. Then the principal square root of A is a positive definite Hermitian diagonal element matrix.*

Proof: See Appendix B.3. \square

Lemma 3.5.5 *Let $A \in \mathbb{C}^{mn \times mn}$ be a diagonal element matrix whose element at position (i, j) is given by*

$$A_{i,j} = \text{diag}(a_1^{(i,j)}, a_2^{(i,j)}, \dots, a_n^{(i,j)})$$

for $i, j = 1, \dots, m$. Then the determinant of A is given by

$$|A| = \prod_{k=1}^n \begin{vmatrix} a_k^{(1,1)} & a_k^{(1,2)} & \dots & a_k^{(1,m)} \\ a_k^{(2,1)} & a_k^{(2,2)} & \dots & a_k^{(2,m)} \\ \vdots & \vdots & \ddots & \vdots \\ a_k^{(m,1)} & a_k^{(m,2)} & \dots & a_k^{(m,m)} \end{vmatrix}.$$

Proof: See Appendix B.4. \square

Using the previous definition and lemmas, we can now prove the following theorem.

Theorem 3.5.2 (Optimal Estimator-Detector Receiver) *There exists estimators $K_i \in \mathbb{C}^{tr \times lr}$ ($i = 0, 1$) such that the corresponding EDR is optimal*

for some decision threshold γ . One such pair of matrices can be found by solving the equation

$$I_{rl} - (I_r \otimes X_i) K_i = \sigma \Lambda_i^{-1/2} \quad (3.50)$$

where $\Lambda_i = \sigma^2 I_{lr} + P_r \otimes X_i P_t X_i^*$ and where the principal square root is taken. This equation always has a solution and γ is given by $|\Lambda_0|^{\sigma^2} |\Lambda_1|^{-\sigma^2}$.

Proof: See Appendix B.5. □

The optimal estimator-detector receiver theorem states that optimality can always be preserved by the use of an EDR. Nevertheless, the equation that must be solved in order to give the analytical expression of the MCEP estimator is rather complicated since it involves taking the principal square root of a matrix whose dimensions increase with the number of transmit and receive antennas. In Section 3.5.3, we will investigate the characteristics of such an estimator under i.i.d. fading by giving its CEP and comparing it to the channel estimators presented in Section 3.3.3.

We show in the following figures the codeword error rates (CER) obtained by simulation using ZF, MMSE and MCEP channel estimators. The parameters of these simulations are: $l = 5$, $t = 3$ and $r = 2$. The codewords X_0 and X_1 are chosen arbitrarily among the non-unitary codewords of size $l \times t$. The transmit and receive covariance matrices P_t and P_r are positive definite Hermitian matrices chosen arbitrarily. X_0 , X_1 , P_t and P_r are chosen only once. We plot the CER under i.i.d. and both transmit and receive correlated fading in Figures 3.5 and 3.6 respectively. Note that the SNR is in logarithmic scale for better readability only.

By construction, the MCEP estimator always provides the best results since it was specifically designed to perform optimal detection. The difference in CER performance between the different estimators is however relatively small. This tends to show that in our framework, a good channel estimate is not crucial in minimising the CER as was previously noticed in the case

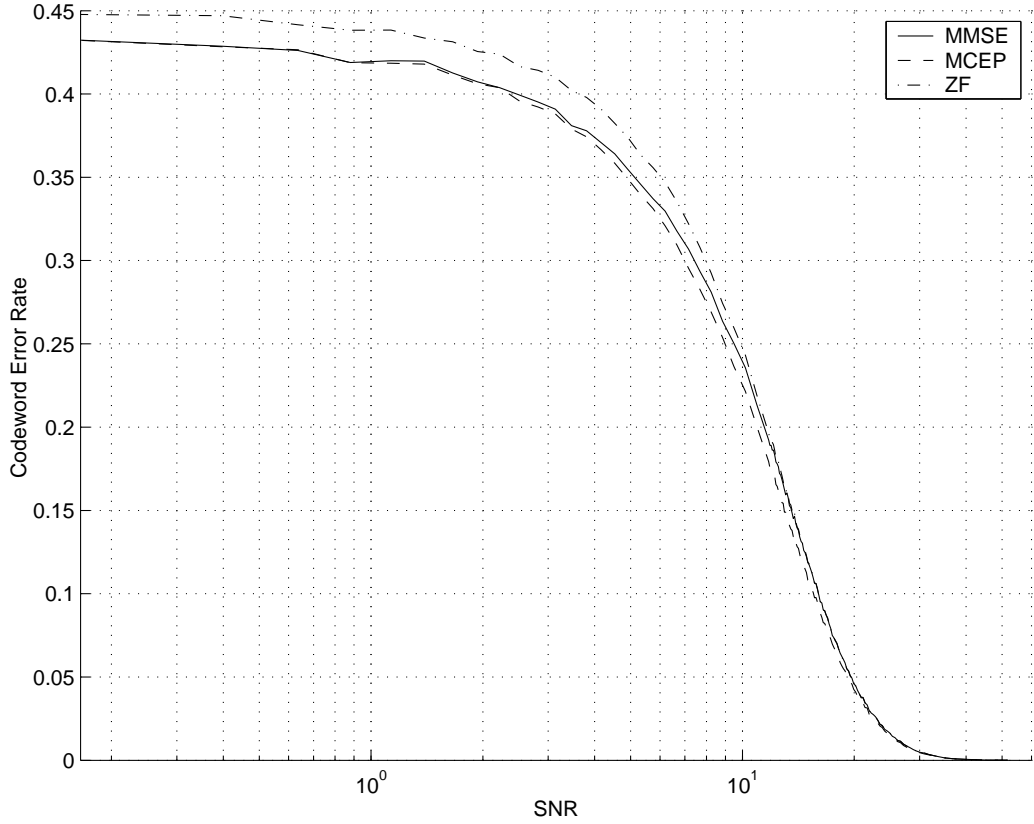


Figure 3.5: Estimation-detection CER for the ZF, MMSE and MCEP estimators in i.i.d. fading.

of unitary codewords. Since computing MCEP estimates using (3.50) can be relatively costly, the practical utility of such an estimator is questionable, especially considering the associated CER performance. Nevertheless, such an estimator is more fundamental from a conceptual point of view since it allows optimal non-coherent ML decoding to be performed in a coherent way.

We also provide in Figures 3.7 and 3.8 a comparison between the channel estimation mean square error obtained by the ZF, MMSE and MCEP estimators under i.i.d. and both transmit and receive correlated fading respectively. Note that the ZF channel estimation MSE is only partially represented since it is not upper-bounded. Based on Figures 3.7 and 3.8, we notice that the MCEP channel estimator does not provide the most accurate channel

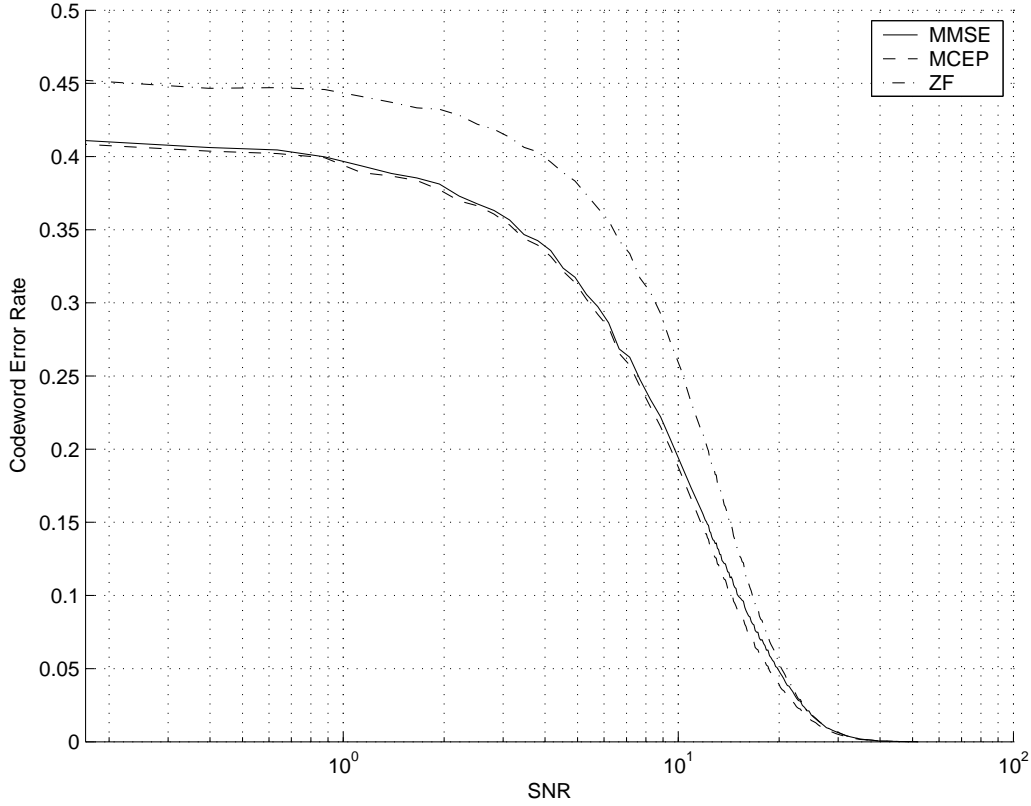


Figure 3.6: Estimation-detection CER for the ZF, MMSE and MCEP estimators in both transmit and receive correlated fading.

estimate while still providing optimal detection. Furthermore, the MCEP estimator provides worse channel estimation than the ZF estimator beyond a specific SNR. This shows that minimising the channel estimation mean square error is not necessarily the best strategy to adopt in order to minimise the decoding error rate.

Now, looking at the proof of Theorem 3.5.2, the uniqueness of such an estimator can be discussed. First, we restrict ourselves to taking the principal square root of a matrix in Equation (3.50) whereas any column-oriented Cholesky decomposition of the matrix Λ_i could lead to an estimator that preserves optimal detection. Nevertheless, this restriction gives us a useful method to prove that such an estimator exists. Second, we look for an es-

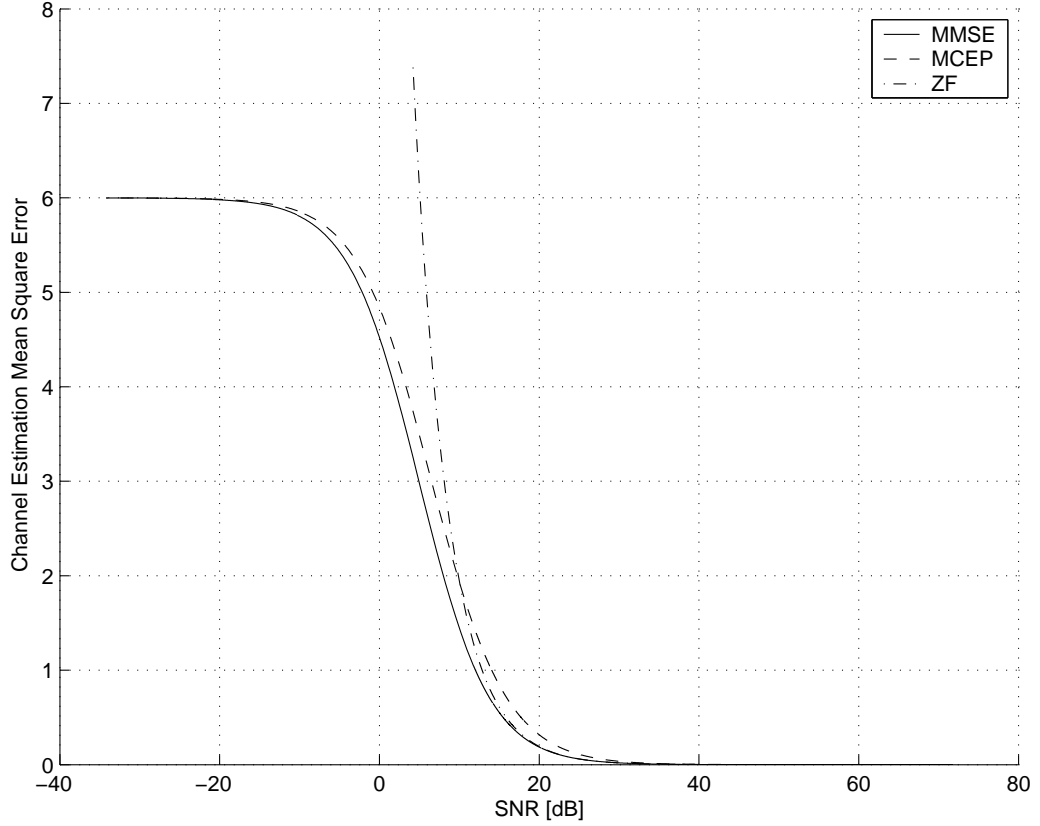


Figure 3.7: ZF, MMSE and MCEP channel estimation mean square error in i.i.d. fading.

timator of the specific form $K_i = V_i D_i U_i^*$ where U_i and V_i are the unitary matrices of the SVD of X_i . Finally, if the codeword X_i is not full rank, some coefficients of the diagonal element matrix D_i can be freely chosen since they are multiplied by the corresponding zero singular values of X_i . This can be intuitively understood by the fact that the channel estimation error (which can be considered as noise) corresponding to directions orthogonal to the vector space spanned by the column of X_i is, under our Gaussian assumptions, completely irrelevant to the decoding process [16].

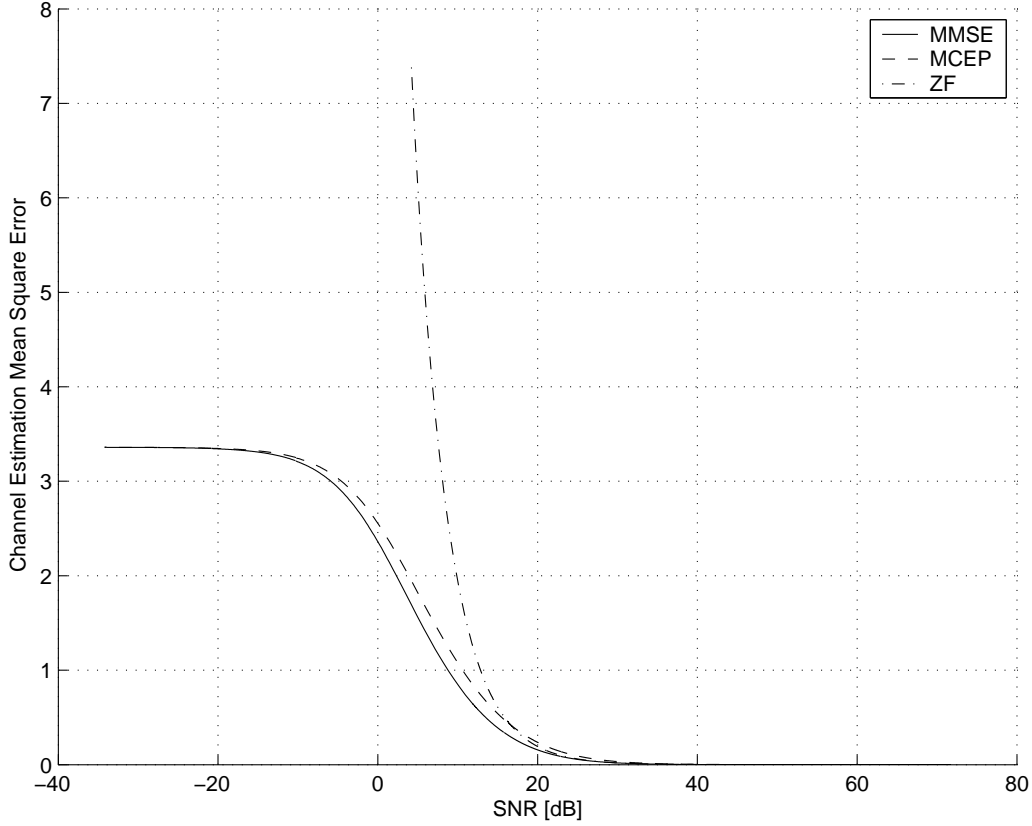


Figure 3.8: ZF, MMSE and MCEP channel estimation mean square error in both transmit and receive correlated fading.

3.5.3 Optimal Estimator in I.I.D. Fading

A closed-form representation of the MCEP estimator in correlated fading is difficult to obtain as seen in the previous section. In the i.i.d. case however, it is possible to obtain the CEP of this estimator as shown in the following theorem.

Theorem 3.5.3 *Let the channel fading be i.i.d. and the singular values of the codeword $X_i \in \mathbb{C}^{l \times t}$ be given by $\sigma_{i,1}, \sigma_{i,2}, \dots, \sigma_{i,t}$. The channel estimator parameters of the MCEP estimators $\tilde{K}_i \in \mathbb{C}^{t \times l}$ ($i = 0, 1$) are given by*

$$\tilde{d}_{i,j} = \begin{cases} \frac{\sigma_{i,j}^*}{|\sigma_{i,j}|^2} \left(1 - \sqrt{\frac{\sigma^2}{|\sigma_{i,j}|^2 + \sigma^2}} \right) & \text{if } \sigma_{i,j} \neq 0 \\ 0 & \text{if } \sigma_{i,j} = 0 \end{cases} \quad (3.51)$$

for $j = 1, \dots, t$.

Proof: See Appendix B.6 □

The CEP obtained for the MCEP estimator in the case of i.i.d. fading allows us to compare it with the ZF and MMSE estimators whose CEP were derived in Section 3.3.3. We can thus analytically see the differences between these three estimators. Following (3.32), (3.37) and (3.51), the corresponding CEP can be written as

$$\tilde{d}_{(i,j),ZF} = \frac{\sigma_{i,j}^*}{|\sigma_{i,j}|^2} \alpha_{ZF} \quad (3.52)$$

$$\tilde{d}_{(i,j),MMSE} = \frac{\sigma_{i,j}^*}{|\sigma_{i,j}|^2} \alpha_{MMSE} \quad (3.53)$$

$$\tilde{d}_{(i,j),MCEP} = \frac{\sigma_{i,j}^*}{|\sigma_{i,j}|^2} \alpha_{MCEP} \quad (3.54)$$

where

$$0 \leq \alpha_{MMSE} = \left(1 - \frac{\sigma^2}{|\sigma_{i,j}|^2 + \sigma^2}\right) \quad (3.55)$$

$$\leq \alpha_{MCEP} = \left(1 - \sqrt{\frac{\sigma^2}{|\sigma_{i,j}|^2 + \sigma^2}}\right) \quad (3.56)$$

$$\leq \alpha_{ZF} = 1. \quad (3.57)$$

If we look at how much the power of the noise is taken into account compared to the signal itself in these three channel estimators, the above equations show that the behaviour of the MCEP channel estimator is in between the MMSE and the ZF. More generally, these three estimators can be seen as elements of a class of estimators whose CEP are of the form

$$\tilde{d}_{(i,j),n} = \frac{\sigma_{i,j}^*}{|\sigma_{i,j}|^2} \left(1 - \left(\frac{\sigma^2}{|\sigma_{i,j}|^2 + \sigma^2}\right)^{1/n}\right) \quad (3.58)$$

where n is some positive parameter set to 1 for MMSE, 2 for MCEP and that tends to 0 for ZF. In this subclass of CEP, referred to as the n -subclass, ZF appears as a limiting case. We show in Figure 3.9 the channel estimation mean square error obtained by estimators corresponding to different values of n . Here $l = 5$, $t = 3$, $r = 2$ and some arbitrary full rank training sequence of

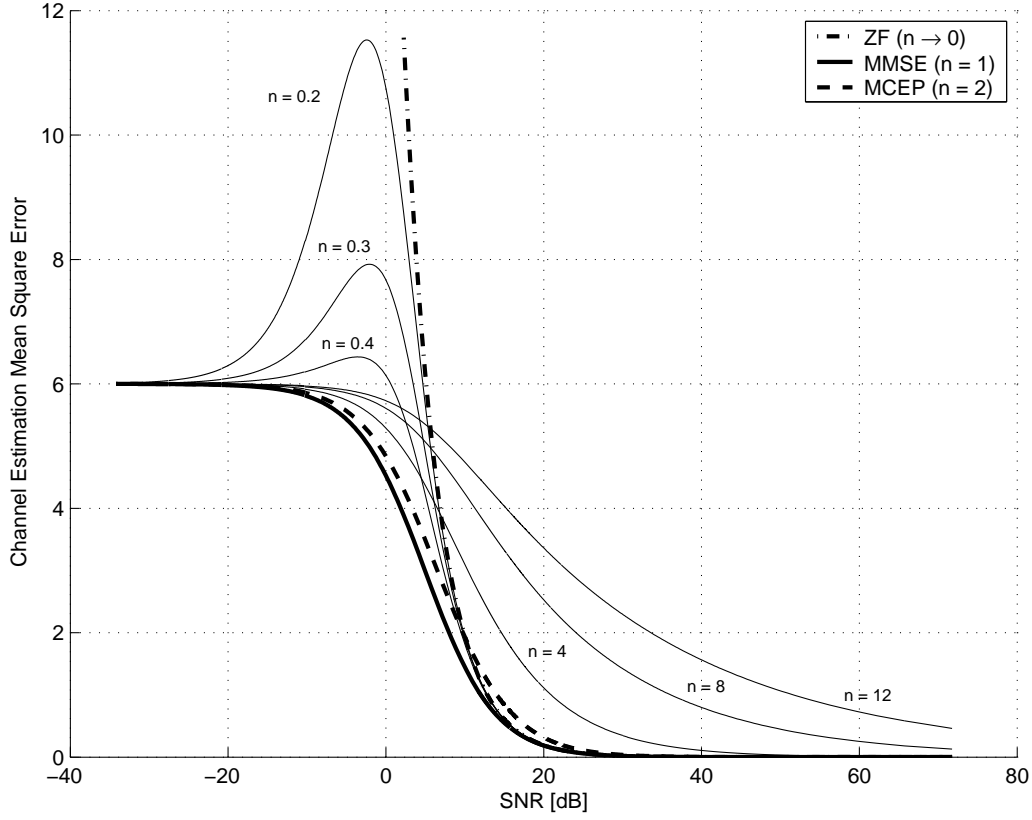


Figure 3.9: Channel estimation mean square error for different values of n .

size $l \times t$ is used. Note that the ZF channel estimation MSE is only partially represented since not upper-bounded. Once more, we see in Figure 3.9 that the MCEP estimator does not minimise the channel estimation MSE while still retaining optimality. The n -subclass also introduces new estimators whose performance would be interesting to evaluate in other applications such as per-survivor processing (PSP) techniques [42, 43]. Furthermore, it can be shown that for $0 < n < 1/2$, the corresponding MSE curves have a global maxima as seen on Figure 3.9 for $n \in \{0.2, 0.3, 0.4\}$, i.e., that at low SNR, the channel estimation MSE increases as the SNR increases. This unexpected behaviour may come from the fact that those estimators are a priori not associated with any meaningful optimisation criterion and thus cannot be expected to behave like ZF or MMSE estimators.

Chapter 4

Conclusion

The aim of this thesis has been to compare coherent and non-coherent decoding schemes in the context of space-time block codes. We have in particular shown that ML non-coherent decoding can actually be decomposed in a coherent way where channel estimates are first computed and then used in a detection step. We have related this approach to the GLRT whose optimality under both the assumptions of unitary codewords and i.i.d. fading was extended using MMSE channel estimates. When no specific assumption is made either on the codewords or on the fading, we have described an estimator that retains optimal decoding. We have also compared the channel estimation MSE and the CER performance of estimator-detector receivers that use different channel estimators. Under i.i.d. fading, we have given the CEP of the channel estimator that preserves optimal detection and proved that in the general case, this estimator does not correspond to an MMSE estimate of the channel. Therefore, we have shown that trying to find a channel estimate as close as possible to the actual channel realisation is not necessarily the best strategy to adopt in order to minimise the codeword error rate.

Since we have derived the expression for the channel estimator that ensures optimal decoding, it would be interesting as future work to investigate the performance of the corresponding channel estimates when applied in the context of PSP techniques. Also, the n-subclass introduced in this thesis leads to a number of channel estimators whose characteristics would be interesting

to investigate. Finally, comparing coherent and non-coherent decoding leads to far more fundamental questions that are directly related with the way a communication system is modelled. Among these, investigating the effect of channel estimation on optimal code design or getting a deep conceptual understanding of the channel estimation process are of particular interest.

Appendix A

Matrix Operations

We present here a few definitions, properties and lemmas useful for the derivation of some results of this thesis. The corresponding proofs can be found in [37] or [38].

A.1 Trace

Definition A.1.1 (Trace) *Let $A \in \mathbb{C}^{n \times n}$. The trace of A , denoted $\text{tr}(A)$, is defined as the sum of its diagonal elements, i.e.*

$$\text{tr}(A) = \sum_{i=1}^n a_{i,i}.$$

Properties A.1.1 (Trace) *Let $A, B, C \in \mathbb{C}^{n \times n}$, $P \in \mathbb{C}^{n \times n}$ a random matrix, k a scalar and $\lambda_1, \lambda_2, \dots, \lambda_n$ the eigenvalues of A . We have the following properties:*

- $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$
- $\text{tr}(kA) = k \text{tr}(A)$
- $\text{tr}(AB) = \text{tr}(BA)$
- $\text{tr}(A^T) = \text{tr}(A)$ and $\text{tr}(A^*) = (\text{tr}(A))^*$
- $\text{tr}(A) = \sum_{i=1}^n \lambda_i$
- $\text{E}[\text{tr}(P)] = \text{tr}(\text{E}[P])$ (expectation and trace commute)

A.2 Kronecker Product and the vec Operator

Definition A.2.1 (Kronecker Product) Let $A \in \mathbb{C}^{n \times p}$ and $B \in \mathbb{C}^{m \times q}$. The Kronecker product of A and B , denoted $A \otimes B$, is the $mn \times pq$ complex matrix defined as

$$A \otimes B = \begin{bmatrix} a_{1,1}B & a_{1,2}B & \dots & a_{1,p}B \\ a_{2,1}B & a_{2,2}B & \dots & a_{2,p}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}B & a_{n,2}B & \dots & a_{n,p}B \end{bmatrix}.$$

The Kronecker product is also called the direct product or the tensor product.

Definition A.2.2 (vec operator) The vec operator creates a column vector of size $mn \times 1$ from a matrix $A \in \mathbb{C}^{m \times n}$ by stacking the $m \times 1$ column vectors of $A = [a_1 a_2 \dots a_n]$ below one another:

$$\text{vec}(A) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

Properties A.2.1 Let A , B and C be complex matrices and a and b be two scalars. We have the following properties:

- $A \otimes (B \otimes C) = (A \otimes B) \otimes C$
- $A \otimes (B + C) = (A \otimes B) + (A \otimes C)$ (for conforming matrices)
- $(A + B) \otimes C = (A \otimes C) + (B \otimes C)$ (for conforming matrices)
- $a \otimes A = A \otimes a = aA$
- $aA \otimes bB = ab(A \otimes B)$
- $(A \otimes B)(C \otimes D) = AC \otimes BD$ (for conforming matrices)
- $(A \otimes B)^* = A^* \otimes B^*$
- $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ (for square non-singular matrices)

- $|A \otimes B| = |A|^n |B|^m$ (for $A \in \mathbb{C}^{m \times m}$ and $B \in \mathbb{C}^{n \times n}$)
- $\text{tr}(A \otimes B) = \text{tr}(A) \text{tr}(B)$
- $\text{vec}(AB) = (I_p \otimes A) \text{vec}(B)$ (for $A \in \mathbb{C}^{m \times n}$ and $B \in \mathbb{C}^{n \times p}$)
- $\text{tr}(AA^*) = \text{tr}(\text{vec}(A) \text{vec}(A)^*)$

A.3 Block Matrices

Definition A.3.1 (Block Matrix) Let $A \in \mathbb{C}^{p \times q}$ a matrix of the form

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{bmatrix}$$

where $A_{i,j} \in \mathbb{C}^{p_i \times q_j}$ for $i = 1, \dots, m$ and $j = 1, \dots, n$ with $\sum_{i=1}^m p_i = p$ and $\sum_{j=1}^n q_j = q$. A is called a block matrix.

Definition A.3.2 (Block Diagonal Matrix) A block diagonal matrix A or diagonal block matrix A , is a square block matrix where the diagonal blocks $A_{i,i}$ are square matrices and the off-diagonal blocks $A_{i,j}$ ($i \neq j$) are the all zero matrices.

Lemma A.3.1 (Block Matrix Inversion) Let $A \in \mathbb{C}^{n \times n}$ be a block matrix partitioned as

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}.$$

Then,

$$\begin{aligned} A^{-1} &= \begin{bmatrix} F_{1,1}^{-1} & -A_{1,1}^{-1}A_{1,2}F_{2,2}^{-1} \\ -F_{2,2}^{-1}A_{2,1}A_{1,1}^{-1} & F_{2,2}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} A_{1,1}^{-1} + A_{1,1}^{-1}A_{1,2}F_{2,2}^{-1}A_{2,1}A_{1,1}^{-1} & -F_{1,1}^{-1}A_{1,2}A_{2,2}^{-1} \\ -A_{2,2}^{-1}A_{2,1}A_{1,1}^{-1} & A_{2,2}^{-1} + A_{2,2}^{-1}A_{2,1}F_{1,1}^{-1}A_{1,2}A_{2,2}^{-1} \end{bmatrix} \end{aligned}$$

where

$$\begin{aligned} F_{1,1} &= A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{2,1} \\ F_{2,2} &= A_{2,2} - A_{2,1}A_{1,1}^{-1}A_{1,2} \end{aligned}$$

assuming that $A_{1,1}^{-1}$, $A_{2,2}^{-1}$, $F_{1,1}^{-1}$ and $F_{2,2}^{-1}$ exist.

Remark: The block matrix inversion lemma can be applied recursively to any conformably partitioned block matrix.

Lemma A.3.2 (Block Diagonal Matrix Determinant) *Let A be a block diagonal matrix with diagonal blocks $A_{i,i}$ for $i = 1, \dots, k$. Then,*

$$|A| = \prod_{i=1}^k |A_{i,i}|.$$

Appendix B

Proofs

B.1 Proof of Lemma 3.3.1

Using Properties A.2.1, we have

$$\begin{aligned}\text{vec}(H) - \text{vec}(\hat{H}) &= \text{vec}(H) - K \text{vec}(Y) \\ &= \text{vec}(H) - K \text{vec}(XH + N) \\ &= \text{vec}(H) - K (I_r \otimes X) \text{vec}(H) - K \text{vec}(N) \\ &= (I_{tr} - K (I_r \otimes X)) \text{vec}(H) - K \text{vec}(N).\end{aligned}$$

Then using the independence of H and N , the fact that $H \sim \mathcal{CN}(0, P_r \otimes P_t)$ and $N \sim \mathcal{CN}(0, \sigma^2 I_{rl})$ and the trace properties A.1.1 we can write,

$$\begin{aligned}& \mathbb{E} \left[\|\text{vec}(H) - \text{vec}(\hat{H})\|^2 \right] \\ &= \mathbb{E} \left[\text{tr} \left((I_{tr} - K (I_r \otimes X)) \text{vec}(H) - K \text{vec}(N) \right) \right. \\ &\quad \left. \cdot \left((I_{tr} - K (I_r \otimes X)) \text{vec}(H) - K \text{vec}(N) \right)^* \right] \\ &= \text{tr} \left((I_{tr} - K (I_r \otimes X))^* (I_{tr} - K (I_r \otimes X)) \mathbb{E} [\text{vec}(H) \text{vec}(H)^*] \right. \\ &\quad \left. + K \mathbb{E} [\text{vec}(N) \text{vec}(N)^*] K^* \right) \\ &= \text{tr} \left((I_{tr} - K (I_r \otimes X))^* (I_{tr} - K (I_r \otimes X)) (P_r \otimes P_t) + \sigma^2 K K^* \right).\end{aligned}$$

□

B.2 Proof of Theorem 3.5.1

Since fading is i.i.d., we consider the ML decoding rule (3.20) and the channel estimates given by (3.24).

Let the MMSE estimator be given by (see Section 3.3.3)

$$\tilde{K}_i = X_i^* (\sigma^2 I_l + X_i X_i^*)^{-1}$$

for $i = 0, 1$. We have that

$$\begin{aligned} Y - X_i \hat{H}_i &= Y - X_i \tilde{K}_i Y \\ &= (I - X_i \tilde{K}_i) Y \\ &= \sigma^2 (\sigma^2 I_l + X_i X_i^*)^{-1} Y \\ &= \sigma^2 \Omega_i^{-1} Y. \end{aligned}$$

Thus using (3.43) and Properties A.1.1, the decision statistic (3.44) becomes

$$\psi(Y) = \text{etr} \left(\sigma^4 (\Omega_1^{-2} - \Omega_0^{-2}) Y Y^* \right).$$

Using the matrix inversion lemma [34, Sec. 0.7.4] and the fact that $X_i^* X_i = I_t$, we have that

$$\begin{aligned} \Omega_i^{-1} &= (\sigma^2 I_l + X_i X_i^*)^{-1} \\ &= \frac{1}{\sigma^2} I_l - \frac{1}{\sigma^4} X_i \left(\frac{1}{\sigma^2} X_i^* X_i + I_t \right)^{-1} X_i^* \\ &= \frac{1}{\sigma^2} I_l - \frac{1}{\sigma^2(\sigma^2 + 1)} X_i X_i^*. \end{aligned}$$

Similarly,

$$\begin{aligned} \Omega_i^{-2} &= (\Omega_i^{-1})^2 \\ &= \left(\frac{1}{\sigma^2} I_l - \frac{1}{\sigma^2(\sigma^2 + 1)} X_i X_i^* \right)^2 \\ &= \frac{1}{\sigma^4} I_l - \left(\frac{1}{\sigma^4} - \frac{1}{(\sigma^2 + 1)^2} \right) X_i X_i^*. \end{aligned}$$

Furthermore, using the determinant formula [39]

$$|I + AB| = |I + BA|,$$

we obtain

$$\begin{aligned}
 |\Omega_i^{-1}| &= \left| \frac{1}{\sigma^2} I_l - \frac{1}{\sigma^2(\sigma^2 + 1)} X_i X_i^* \right| \\
 &= \left| \frac{1}{\sigma^2} I_t - \frac{1}{\sigma^2(\sigma^2 + 1)} X_i^* X_i \right| \\
 &= \left(\frac{1}{\sigma^2 + 1} \right)^t
 \end{aligned}$$

which does not depend on i . Finally, since $(\Omega_1^{-1} - \Omega_0^{-1})$ and $(\Omega_1^{-2} - \Omega_0^{-2})$ differ only by a positive multiplicative constant and the decision threshold is equal to 1, the two decision rules (3.45) and (3.20) are equivalent with $\gamma = 1$ and thus the EDR is optimal. \square

B.3 Proof of Lemma 3.5.4

Since A is Hermitian, it can be unitarily diagonalised, i.e.

$$A = U D U^*$$

where $U \in \mathbb{C}^{mn \times mn}$ is a unitary matrix and $D \in \mathbb{C}^{mn \times mn}$ is a diagonal matrix with coefficients d_1, d_2, \dots, d_{mn} . Then,

$$a_{s,t} = \sum_{p=1}^{mn} d_p u_{s,p} u_{t,p}^*.$$

Since A is a diagonal element matrix, $a_{s,t} = 0$ for all $(s, t) \in \Upsilon$ where

$$\Upsilon = \{(s, t) \in \{1, \dots, mn\}^2 \text{ such that } |s - t| \neq ln \text{ for } l = 0, \dots, m - 1\}.$$

Also, being Hermitian and positive definite, there exists a unique positive definite Hermitian matrix B such that $B^2 = A$ [34, Th. 7.2.6]. The matrix B is called the principal square root of A and is obtained by taking the (unique) real and positive square root of every eigenvalue. This is the square root we are going to consider in the rest of this proof.

To show that $A^{1/2}$ is a diagonal element matrix, we only have to ensure

that the element at the s -th row and the t -th column of $A^{1/2}$, denoted by $(a^{1/2})_{s,t}$, is also equal to 0 for $(s, t) \in \Upsilon$. Using Lemma 3.5.2, we know it is the case for A^k when $k \in \mathbb{N} \setminus \{0\}$. Since

$$A^k = UD^kU^*$$

where $D^k \in \mathbb{C}^{mn \times mn}$ is a diagonal matrix with coefficients $d_1^k, d_2^k, \dots, d_{mn}^k$, we can write

$$(a^k)_{s,t} = \sum_{p=1}^{mn} d_p^k u_{s,p} u_{t,p}^*$$

where $(a^k)_{s,t}$ denotes the element at the s -th row and the t -th column of A^k . Furthermore, for $s \neq t$,

$$\sum_{p=1}^{mn} u_{s,p} u_{t,p}^* = 0$$

since the rows of U are orthogonal. Thus, for any $(s, t) \in \Upsilon$ we can write the following system of equations

$$\underbrace{\begin{bmatrix} 1 & 1 & \dots & 1 \\ d_1 & d_2 & \dots & d_{mn} \\ d_1^2 & d_2^2 & \dots & d_{mn}^2 \\ \vdots & \vdots & \ddots & \vdots \\ d_1^{mn-1} & d_2^{mn-1} & \dots & d_{mn}^{mn-1} \end{bmatrix}}_M \begin{bmatrix} u_{s,1} u_{t,1}^* \\ u_{s,2} u_{t,2}^* \\ \vdots \\ u_{s,mn} u_{t,mn}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where $M \in \mathbb{C}^{mn \times mn}$ is a Vandermonde matrix whose determinant is [34, Sec. 0.9.11]

$$\prod_{\substack{i,j=1 \\ i < j}}^{mn} (d_i - d_j).$$

If the eigenvalues d_1, d_2, \dots, d_{mn} are distinct, the determinant of M is non zero and the above system of equations only has the all-zero trivial solution. Thus replacing d_1, d_2, \dots, d_{mn} by $d_1^{1/2}, d_2^{1/2}, \dots, d_{mn}^{1/2}$ will give $(a^{1/2})_{s,t} = 0$ for $(s, t) \in \Upsilon$. If there are only r distinct eigenvalues ($r < mn$), we can rewrite the above system of equations by grouping the $u_{s,p} u_{t,p}^*$ corresponding to equal eigenvalues, remove enough rows to get a Vandermonde matrix A of size $r \times r$ and apply the same reasoning as before. \square

B.4 Proof of Lemma 3.5.5

The proof follows directly from the fact that we can swap the rows and columns of A an even number of times to get a block diagonal matrix $\tilde{A} \in \mathbb{C}^{mn \times mn}$ whose diagonal blocks are given by

$$\tilde{A}_{k,k} = \begin{bmatrix} a_k^{(1,1)} & a_k^{(1,2)} & \cdots & a_k^{(1,m)} \\ a_k^{(2,1)} & a_k^{(2,2)} & \cdots & a_k^{(2,m)} \\ \vdots & \vdots & \ddots & \vdots \\ a_k^{(m,1)} & a_k^{(m,2)} & \cdots & a_k^{(m,m)} \end{bmatrix}$$

for $k = 1, \dots, n$ and whose determinant remains unchanged. We conclude the proof using the block diagonal matrix determinant lemma A.3.2. \square

B.5 Proof of Theorem 3.5.2

The proof is divided into two parts. In Part A, we show that a sufficient condition on the estimator K_i for the optimality to be preserved is to satisfy

$$I_{rl} - (I_r \otimes X_i) K_i = k \Lambda_i^{-1/2} \quad (\text{B.1})$$

for $i = 0, 1$ and some positive constant k and where the principal square root is taken. We also give the corresponding decision threshold γ . In Part B, we prove that (B.1) always has a solution and therefore that an estimator preserving optimal detection always exists. We show that in that case, the constant k has to be equal to σ .

PART A: Conditioned upon the codeword X_i ($i = 0, 1$), the channel estimate is given by $\text{vec}(\hat{H}_i) = K_i \text{vec}(Y)$. Using this channel estimate and Properties A.1.1 and A.2.1, the decoding metric (3.42) is

$$\begin{aligned} & \| \text{vec}(Y) - \text{vec}(X_i \hat{H}_i) \|^2 \\ &= \text{tr} \left(\text{vec}(Y - X_i \hat{H}_i) \text{vec}(Y - X_i \hat{H}_i)^* \right) \\ &= \text{tr} \left((I_{rl} - (I_r \otimes X_i) K_i)^* (I_{rl} - (I_r \otimes X_i) K_i) \text{vec}(Y) \text{vec}(Y)^* \right) \end{aligned}$$

and the decision statistic (3.44) becomes

$$\psi(Y) = \text{etr} \left((\Psi_1 - \Psi_0) \text{vec}(Y) \text{vec}(Y)^* \right)$$

where

$$\Psi_i = (I_{rl} - (I_r \otimes X_i) K_i)^* (I_{rl} - (I_r \otimes X_i) K_i).$$

Comparing $\psi(Y)$ to the optimal decision statistic (3.14), we see that finding a matrix K_i such that

$$\Psi_i = \tilde{k} \Lambda_i^{-1}$$

for some positive constant \tilde{k} , allows the estimator-detector receiver of Section 3.4 to be optimal for some threshold γ . The general solution to this problem would be to consider a column-oriented Cholesky decomposition of the positive definite Hermitian matrix Λ_i^{-1} . But in this case, the existence of such a K_i would be difficult to prove. Nevertheless, since Λ_i is a positive definite Hermitian matrix (real and positive eigenvalues), so is Λ_i^{-1} . Thus, there exists a unique positive definite Hermitian matrix A_i such that $A_i^* A_i = A_i^2 = \Lambda_i^{-1}$ [34, Th. 7.2.6]. This matrix can be obtained by considering the eigenvalue decomposition of Λ_i^{-1} and taking the principal square root of every eigenvalue. Therefore, a sufficient condition is to find a matrix K_i such that

$$I_{rl} - (I_r \otimes X_i) K_i = k \Lambda_i^{-1/2}$$

for some positive constant k . In other words, we want to show that this matrix A_i can be written of the form

$$I_{rl} - (I_r \otimes X_i) K_i$$

for some matrix K_i . Note that for (3.13) to be equivalent to (3.45), the threshold γ must be equal to $|\Lambda_0|^{k^2} |\Lambda_1|^{-k^2}$.

PART B: We now have to show that there exists a matrix K_i such that equation (B.1) is satisfied for some positive constant k . Without loss of generality, we can consider the case where $P_t = I_t$ (receive correlation only). Otherwise, we can simply replace X_i by $X_i P_t^{1/2}$ since no assumption is made

on the codewords.

Let $X_i = U_i \Sigma_i V_i^*$ be the SVD of X_i where $U_i \in \mathbb{C}^{l \times l}$ and $V_i \in \mathbb{C}^{t \times t}$ are unitary matrices and $\Sigma_i \in \mathbb{C}^{l \times t}$ is of the form

$$\Sigma_i = \begin{bmatrix} \bar{\Sigma}_i \\ O_{(l-t) \times t} \end{bmatrix}$$

with $\bar{\Sigma}_i = \text{diag}(\sigma_{i,1}, \sigma_{i,2}, \dots, \sigma_{i,t})$.

Using Properties A.2.1, we have

$$\begin{aligned} \Lambda_i &\triangleq \sigma^2 I_{rl} + P_r \otimes X_i X_i^* \\ &= \sigma^2 (I_r \otimes I_l) + P_r \otimes U_i \Sigma_i \Sigma_i^* U_i^* \\ &= \sigma^2 (I_r \otimes U_i U_i^*) + (I_r \otimes U_i) (P_r \otimes I_l) (I_r \otimes \Sigma_i \Sigma_i^*) (I_r \otimes U_i^*) \\ &= (I_r \otimes U_i) \sigma^2 I_{rl} (I_r \otimes U_i^*) + (I_r \otimes U_i) (P_r \otimes I_l) (I_r \otimes \Sigma_i \Sigma_i^*) (I_r \otimes U_i^*) \\ &= (I_r \otimes U_i) (\sigma^2 I_{rl} + (P_r \otimes I_l) (I_r \otimes \Sigma_i \Sigma_i^*)) (I_r \otimes U_i^*) \\ &= (I_r \otimes U_i) (\sigma^2 I_{rl} + (P_r \otimes \Sigma_i \Sigma_i^*)) (I_r \otimes U_i^*) \end{aligned}$$

and

$$\Lambda_i^{-1/2} = (I_r \otimes U_i) \underbrace{(\sigma^2 I_{rl} + (P_r \otimes \Sigma_i \Sigma_i^*))}_{Q_i}^{-1/2} (I_r \otimes U_i^*).$$

Note that the above equality can easily be verified by squaring and taking the inverse of $\Lambda_i^{-1/2}$ to obtain Λ_i . Furthermore,

$$I_{rl} - (I_r \otimes X_i) K_i = (I_r \otimes U_i) (I_r \otimes U_i^*) - (I_r \otimes U_i) (I_r \otimes \Sigma_i) (I_r \otimes V_i^*) K_i.$$

Thus if we look for a matrix K_i of the form

$$K_i = (I_r \otimes V_i) D_i (I_r \otimes U_i^*)$$

for some matrix $D_i \in \mathbb{C}^{rt \times rl}$, we obtain

$$I_{rl} - (I_r \otimes X_i) K_i = (I_r \otimes U_i) \underbrace{(I_{rl} - (I_r \otimes \Sigma_i) D_i)}_{R_i} (I_r \otimes U_i^*).$$

Equation (B.1) is satisfied if

$$(I_{rl} - (I_r \otimes \Sigma_i) D_i) = k (\sigma^2 I_{rl} + (P_r \otimes \Sigma_i \Sigma_i^*))^{-1/2}. \quad (\text{B.2})$$

We will now show that we can always find a matrix D_i and a positive constant k such that (B.2) is satisfied.

Let D_i be written

$$D_i = \begin{bmatrix} D_{i1,1} & D_{i1,2} & \cdots & D_{i1,r} \\ D_{i2,1} & D_{i2,2} & \cdots & D_{i2,r} \\ \vdots & \vdots & \ddots & \vdots \\ D_{ir,1} & D_{ir,2} & \cdots & D_{ir,r} \end{bmatrix}$$

where $D_{i_{s,t}} \in \mathbb{C}^{t \times l}$ is of the form

$$D_{i_{s,t}} = \begin{bmatrix} \bar{D}_{i_{s,t}} & O_{t \times (l-t)} \end{bmatrix}$$

with $\bar{D}_{i_{s,t}} \in \mathbb{C}^{t \times t}$ a diagonal matrix and $s, t = 1, \dots, r$.

Then, $R_i \in \mathbb{C}^{rl \times rl}$ is an diagonal element matrix whose element at position (s, t) $R_{i_{s,t}} \in \mathbb{C}^{l \times l}$ is given by

$$R_{i_{s,t}} = \begin{cases} \begin{bmatrix} I_t - \bar{\Sigma}_i \bar{D}_{i_{s,t}} & O_{t \times (l-t)} \\ O_{(l-t) \times t} & I_{(l-t) \times (l-t)} \end{bmatrix} & \text{for } s, t = 1, \dots, r, s = t \\ \begin{bmatrix} -\bar{\Sigma}_i \bar{D}_{i_{s,t}} & O_{t \times (l-t)} \\ O_{(l-t) \times t} & O_{(l-t) \times (l-t)} \end{bmatrix} & \text{for } s, t = 1, \dots, r, s \neq t \end{cases}.$$

Also, $Q_i \in \mathbb{C}^{rl \times rl}$ is a positive definite Hermitian diagonal element matrix whose element at position (s, t) $Q_{i_{s,t}} \in \mathbb{C}^{l \times l}$ is given by

$$Q_{i_{s,t}} = \begin{cases} \begin{bmatrix} \sigma^2 I_t + p_{r,s,t} \bar{\Sigma}_i \bar{\Sigma}_i^* & O_{t \times (l-t)} \\ O_{(l-t) \times t} & \sigma^2 I_{(l-t) \times (l-t)} \end{bmatrix} & \text{for } s, t = 1, \dots, r, s = t \\ \begin{bmatrix} p_{r,s,t} \bar{\Sigma}_i \bar{\Sigma}_i^* & O_{t \times (l-t)} \\ O_{(l-t) \times t} & O_{(l-t) \times (l-t)} \end{bmatrix} & \text{for } s, t = 1, \dots, r, s \neq t \end{cases}$$

where $p_{r,s,t}$ denotes the element at the s -th row and the t -th column of P_r .

Note that if X_i is not full rank (i.e. some singular values are zero), the form of R_i and Q_i does not change. Only the dimensions of the identity and all-zero

matrices inside the diagonal elements change (assuming that the singular values are sorted in decreasing order) and the following discussion can still be applied. Thus, we will consider that X_i is full rank for the rest of the proof.

Since Q_i is a positive definite Hermitian diagonal element matrix using Lemma 3.5.3 and Lemma 3.5.4, so is $Q_i^{-1/2}$ (principal square root). From Lemma 3.5.5, the proof of Lemma 3.5.1 and the block matrix inversion lemma A.3.1, we see that $Q_i^{-1/2}$ has to be a diagonal element matrix whose element at position (s, t) $(Q_i^{-1/2})_{s,t} \in \mathbb{C}^{l \times l}$ is given by

$$(Q_i^{-1/2})_{s,t} = \begin{cases} \begin{bmatrix} (\tilde{Q}_i)_{s,t} & O_{t \times (l-t)} \\ O_{(l-t) \times t} & \sigma^{-1} I_{(l-t) \times (l-t)} \end{bmatrix} & \text{for } s, t = 1, \dots, r, s = t \\ \begin{bmatrix} (\tilde{Q}_i)_{s,t} & O_{t \times (l-t)} \\ O_{(l-t) \times t} & O_{(l-t) \times (l-t)} \end{bmatrix} & \text{for } s, t = 1, \dots, r, s \neq t \end{cases}$$

for some diagonal matrix $(\tilde{Q}_i)_{s,t} \in \mathbb{C}^{t \times t}$. Otherwise $Q_i^{-1/2}$ cannot be a positive definite Hermitian matrix and Q_i of the form given above.

Comparing the form of $Q_i^{-1/2}$ and R_i , we see that solving the equation (B.2) reduces to solve coordinate-wise the system corresponding to the diagonal elements of $(\tilde{Q}_i)_{s,t}$. Since $\sigma_{i,1}, \sigma_{i,2}, \dots, \sigma_{i,t}$ are non-zero, this system always has a solution. Furthermore, k has to be equal to σ . \square

B.6 Proof of Theorem 3.5.3

Under i.i.d. fading, $P_t = I_t$ and $P_r = I_r$. Thus solving Equation (3.50) for an estimator of the form $K_i = I_r \otimes \tilde{K}_i \in \mathbb{C}^{tr \times lr}$ as given by (3.24) reduces to solving

$$I_l - X_i \tilde{K}_i = \sigma(\sigma^2 I_l + X_i X_i^*)^{-1/2}.$$

Let $X_i = U_i \Sigma_i V_i^*$ be the SVD of X_i and $\tilde{K}_i = V_i \tilde{D}_i U_i^*$. The above equation reduces to

$$I_l - U_i \Sigma_i \tilde{D}_i U_i^* = U_i (I_l - \Sigma_i \tilde{D}_i) U_i^* = \sigma U_i (\sigma^2 I_l + \Sigma_i \Sigma_i^*)^{-1/2} U_i^*.$$

Note that $\Sigma_i \tilde{D}_i$ is of the form

$$\begin{aligned} & \begin{bmatrix} \bar{\Sigma}_i \\ O_{(l-t) \times t} \end{bmatrix} \begin{bmatrix} \bar{D}_i & O_{t \times (l-t)} \end{bmatrix} \\ &= \begin{bmatrix} \bar{\Sigma}_i \bar{D}_i & O_{t \times (l-t)} \\ O_{(l-t) \times t} & O_{(l-t) \times (l-t)} \end{bmatrix} \end{aligned}$$

where

$$\begin{aligned} \bar{\Sigma}_i &= \text{diag}(\sigma_{i,1}, \sigma_{i,2}, \dots, \sigma_{i,t}) \\ \bar{D}_i &= \text{diag}(\tilde{d}_{i,1}, \tilde{d}_{i,2}, \dots, \tilde{d}_{i,t}). \end{aligned}$$

Coordinate-wise we obtain

$$1 - \sigma_{i,j} \tilde{d}_{i,j} = \sigma (|\sigma_{i,j}|^2 + \sigma^2)^{-1/2}$$

or

$$\tilde{d}_{i,j} = \begin{cases} \frac{\sigma_{i,j}^*}{|\sigma_{i,j}|^2} \left(1 - \sqrt{\frac{\sigma^2}{|\sigma_{i,j}|^2 + \sigma^2}} \right) & \text{if } \sigma_{i,j} \neq 0 \\ 0 & \text{if } \sigma_{i,j} = 0 \end{cases}$$

for $j = 1, \dots, t$. Note that when $\sigma_{i,j} = 0$, the corresponding $\tilde{d}_{i,j}$ can be freely chosen. Here we arbitrarily set its value to 0. \square

Bibliography

- [1] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–159, August 1996.
- [2] G. J. Foschini and M. J. Gans, “On limits of wireless communications in a fading environment when using multiple antennas,” *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, March 1998.
- [3] V. Tarokh, N. Seshadri and A. R. Calderbank, “Space-time codes for high data rate wireless communication: Performance criterion and code construction,” *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, March 1998.
- [4] S. M. Alamouti, “A simple transmit diversity technique for wireless communications,” *IEEE Journal of Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, October 1998.
- [5] Ĭ. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Transactions on Telecommunication*, vol. 10, no. 6, pp. 585–595, November/December 1999. Originally published as Technical Memorandum, Bell Laboratories, Lucent Technologies, October 1995.
- [6] T. L. Marzetta and B. M. Hochwald, “Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading,” *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, January 1999.
- [7] B. M. Hochwald and T. L. Marzetta, “Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading,” *IEEE*

- Transactions on Information Theory*, vol. 46, no. 2, pp. 543–564, March 2000.
- [8] B. L. Hughes, “Differential space-time modulation,” *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2567–2578, November 2000.
- [9] L. Zheng and D. N. C. Tse, “Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel,” *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, February 2002.
- [10] V. Tarokh, A. Naguib, N. Seshadri and A. R. Calderbank, “Space-time codes for high data rate wireless communication: Performance criteria in the presence of channel estimation errors, mobility and multiple paths,” *IEEE Transactions on Communications*, vol. 47, no. 2, pp. 199–207, February 1999.
- [11] A. Scaglione, P. Stoica, S. Barbarossa, G. Giannakis and H. Sampath, “Optimal designs for space time linear precoders and decoders,” *IEEE Transactions on Signal Processing*, vol. 50, no. 5, pp. 1051–1064, May 2002.
- [12] H. L. van Trees, *Detection, estimation and modulation theory. Part I*. John Wiley and Sons, New-York, 1968.
- [13] J. H. Winters, “On the capacity of radio communication systems with diversity in a Rayleigh fading environment,” *IEEE Journal of Selected Areas in Communications*, vol. 5, no. 5, pp. 871–878, June 1987.
- [14] J. Zander, “Performance of optimum transmitter power control in cellular radio systems,” *IEEE Transactions on Vehicular Technology*, vol. 41, pp. 57–62, February 1992.
- [15] B. Sklar, “Rayleigh fading channels in mobile digital communication systems. Part I: Characterization,” *IEEE Communications Magazine*, vol. 35, pp. 90–100, July 1997.

-
- [16] J. G. Proakis, *Digital communications*. McGraw-Hill, New-York, 1989.
 - [17] A. J. Paulraj, D. Gesbert and C. Papadias, *Smart antennas for mobile communications*. Encyclopedia for Electrical Engineering, John Wiley Publishing Co., 2000.
 - [18] A. Goldsmith, S. A. Jafar, N. Jindal and S. Vishwanath, “Capacity limits of MIMO channels,” *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 684–702, June 2003.
 - [19] C. E. Shannon, “A mathematical theory of communications,” *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and October 1948.
 - [20] R. G. Gallager, *Information theory and reliable communications*. John Wiley and Sons, New-York, 1968.
 - [21] J. K. Cavers, “An analysis of pilot symbol assisted modulation for Rayleigh flat fading channels,” *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 686–693, November 1991.
 - [22] S. Zhou and G. B. Giannakis, “How accurate channel prediction needs to be for adaptive modulation in Rayleigh MIMO channels?,” *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 81–84, April 6-10, 2003.
 - [23] R. R. Müller, “A random matrix model of communication via antenna arrays,” *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2495–2506, September 2002.
 - [24] L. Hanlen and M. Fu, “Multiple antenna wireless communication systems: Limits to capacity growth,” *Wireless Communications and Networking Conference*, vol. 1, pp. 172–176, March 17-21, 2002.
 - [25] S. A. Jafar and A. Goldsmith, “Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information,” *IEEE International Symposium on Information Theory*, p. 470, June 29 - July 4, 2003.

-
- [26] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Transactions on Information Theory*, vol. 28, pp. 55–67, January 1982.
 - [27] V. Tarokh, H. Jafarkhani and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.
 - [28] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *IEEE Proceedings*, vol. 86, pp. 1951–1968, October 1998.
 - [29] A. N. D'Andrea, A. Diglio and U. Mengali "Symbol-aided channel estimation with nonselective Rayleigh fading channels," *IEEE Transactions on Vehicular Technology*, vol. 44, pp. 41–49, February 1995.
 - [30] Y. Akyildiz and B. D. Rao, "Maximum ratio combining performance with imperfect channel estimates," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, pp. 2485–2488, May 13-17, 2002.
 - [31] G. E. Øien, H. Holm and K. J. Hole, "Adaptive coded modulation with imperfect channel state information: System design and performance analysis aspects," *Proceedings of IEEE International Symposium on Advances in Wireless Communications*, Victoria, BC, Canada, September 23-24, 2002.
 - [32] S. Da-Shan, G. J. Foschini, M. J. Gans and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Transactions on Communications*, vol. 48, pp. 502–513, March 2000.
 - [33] D. Chishik, J. Ling, P. Wolniansky and R. Valenzuela, "Multiple input multiple output measurements and modeling in Manhattan," *Proceedings of IEEE Vehicular Technology Conference*, vol. 1, pp. 107–110, September 24-28, 2002.

-
- [34] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, Cambridge, U.K., 1996.
 - [35] J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson and P. O. Börjesson, "On channel estimation in OFDM systems," *IEEE Conference on Vehicular Technology*, vol. 2, pp. 815–819, July 25–28, 1995.
 - [36] J. H. Kotecha and A. K. Sayeed "Optimal signal design for estimation of correlated MIMO channels," *IEEE International Conference on Communications*, vol. 5, pp. 3170–3174, May 11–15, 2003.
 - [37] A. Graham, *Kronecker products and matrix calculus with applications*. Ellis Horwood, Chichester, 1981.
 - [38] F. A. Graybill, *Matrices with applications in statistics*. Wadsworth, Belmont, California, 1983.
 - [39] A. K. Gupta and D. K. Nagar, *Matrix variate distributions*. Monographs and Surveys in Pure and Applied Mathematics, Chapman & Hall/CRC, Boca Raton, 2000.
 - [40] J. W. Carlyle and J. B. Thomas, "On nonparametric signal detectors," *IEEE Transactions on Information Theory*, vol. 10, pp. 146–152, April 1964.
 - [41] M. V. Eyuboğlu and S. U. H. Qureshi, "Reduced-state sequence estimation for coded modulation of intersymbol interference channels," *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 989–995, August 1989.
 - [42] R. Raheli, A. Polydoros and C. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environment," *IEEE Transactions on Communications*, vol. 43, pp. 354–364, February/March/April 1995.
 - [43] G. Auer, G. J. R. Povey and D. I. Laurenson, "Per-survivor processing applied to decision directed channel estimation for a coherent diversity

- receiver,” *IEEE 5th International Symposium on Spread Spectrum Techniques and Applications*, vol. 2, pp. 580–584, September 2-4, 1998.
- [44] G. B. Giannakis and S. Zhou, “Optimal transmit-diversity precoders for random fading channels,” *IEEE Global Telecommunications Conference*, vol. 3, pp. 1839–1843, November 27 - December 1, 2000.
- [45] M. Brehler and M. K. Varanasi, “Asymptotic error probability analysis of quadratic receivers in Rayleigh-fading channels with applications to a unified analysis of coherent and noncoherent space-time receivers,” *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2383–2399, September 2001.
- [46] V. Erceg, P. Soma, D. S. Baum and A. J. Paulraj, “Capacity obtained from multi-input multi-output channel measurements in fixed wireless environments at 2.5 GHz,” *IEEE International Conference on Communications*, vol. 1, pp. 396–400, April 28 - May 2, 2002.
- [47] E. Biglieri and G. Taricco, “Exact pairwise error probability of space-time codes,” *IEEE International Conference on Communications*, vol. 3, pp. 1373–1376, April 28 - May 2, 2002.
- [48] E. Biglieri, G. Taricco and A. Tulino, “Linear receivers for multiple-antenna communication channels: An asymptotic analysis,” *IEEE International Conference on Communications*, vol. 3, pp. 1944–1948, April 28 - May 2, 2002.
- [49] S. Zhou and G. B. Giannakis, “Optimal transmitter eigen-beamforming and space-time block coding based on channel correlations,” *IEEE Transactions on Information Theory*, vol. 49, no. 7, July 2003.
- [50] H. Lee, S. Patil and R. G. Raj, *Fundamental overview and simulation of MIMO systems for space-time coding and spatial multiplexing*. Wireless Networking and Communications Group, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas, USA.